Science Education

SEPTEMBER 2015 / VOLUME 99 / ISSUE NO 5

WILEY

Science Education

EDITOR
John L. Rudolph
Science Education
University of Wisconsin–Madison
Madison, WI 53706

PAST EDITORS
Gregory J. Kelly (2006–2012)
Nancy W. Brickhouse (2001–2006)
Richard A. Duschl (1993–2001)
Leopold E. Klopfer (1979–1992)

Assistant to the Editor: Shannon Vakil

SCIENCE EDUCATION (ISSN 0036-8326 (Print); ISSN 1098-237X (Online)) is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.
Postmaster: Send all address changes to SCIENCE EDUCATION, Journal Customer Services, John Wiley & Sons Inc., C/O The Sheridan Press, PO Box 465, Hanover, PA 17331.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts. Follow our progress at www.wiley.com/go/citizenship
Information for subscribers: SCIENCE EDUCATION is published in 6 issues per year. Institutional subscription prices for 2015 are: Print & Online: US$2331 (US), US$2457 (Rest of World), €1602 (Europe), £1269 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST/HST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2011, where available. For other pricing options, including access information and terms and conditions, please visit wileyonline library.com/access. **Delivery Terms and Legal Title:** Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability. **Journal Customer Services:** For ordering information, claims and any enquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.
Americas: Email: cs-journals@wiley.com; Tel: 1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada). **Europe, Middle East and Africa:** Email: cs-journals@wiley.com; Tel: 44 (0) 1865 778315. **Asia Pacific:** Email: cs-journals@wiley.com; Tel: 65 6511 8000. **Japan:** For Japanese speaking support, Email: cs-japan@wiley.com; Tel: 65 6511 8010 or Tel (toll-free) 005 316 50 480. **Visit our Online Customer Get-Help** available in 7 languages at www.wileycustomerhelp.com/ask
Back issues: Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue—Ste 300, Hudson, NY 12534, USA. Tel: 1 518 822-9300, Fax: 1 518 822-9305, Email: psc@periodicals.com
Advertising Sales: Inquiries concerning advertising should be forwarded to Advertising Sales Manager, Joe Tomaszewski, John Wiley & Sons, 111 River Street, Hoboken, NJ 07030: (201) 748-8895. Email: jtomaszews@wiley.com
Production Editor: Anju Upadhyay (anupadhyay@wiley.com).
Other correspondence: All other correspondence should be addressed to Science Education, Publisher, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030.
Information: For submission instructions, subscription, and all other information and to view this journal online, visit: wileyonlinelibrary.com/sce
Online Open: SCIENCE EDUCATION accepts articles for Open Access publication. Please visit http://olabout.wiley.com/WileyCDA/Section/id-406241.html for further information about OnlineOpen.

This journal is indexed by *CIJE: Current Index to Journals in Education (ERIC), Contents Pages in Education (T&F), Current Abstracts (EBSCO), Current Contents® /Social & Behavioral Sciences (Thomson ISI), Education Index/Abstracts (HW Wilson), Educational Research Abstracts Online (T&F), ERIC Database (Education Resources Information Center), FRANCIS Database (INIST/CNRS), IBR & IBZ: International Bibliographies of Period-ical Literature (KG Saur), Informal Science.org (University of Pittsburgh Center for Learning in Out-of-School Environments), Journal Citation Reports/Social Science Edition (Thomson ISI), Psychological Abstracts/PsycINFO (APA), Research into Higher Education Abstracts (T&F), SCOPUS (Elsevier), Social Sciences Citation Index® (Thomson ISI), Social SciSearch® (Thomson ISI), and Web of Science® (Thomson ISI).* Printed in the USA by The Sheridan Group.

♾ **This paper meets the requirements of ANI/NISO Z39.48-1992 (Permanence of paper).**

Science
Education

CONTENTS

BOARD OF REVIEWERS

Science
Education

Toward a Descriptive Science of Teaching: How the TDOP Illuminates the Multidimensional Nature of Active Learning in Postsecondary Classrooms

MATTHEW T. HORA
Department of Liberal Arts and Applied Studies and the Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, WI 53715, USA

Received 22 October 2013; accepted 17 March 2015
DOI 10.1002/sce.21175
Published online 10 June 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Detailed accounts of teaching can shed light on the nature and prevalence of active learning, yet common approaches reduce teaching to unidimensional descriptors or binary categorizations. In this paper, I use the instructional systems-of-practice framework and the Teaching Dimensions Observation Protocol (TDOP) to advance an approach to thinking about teaching in science classrooms in more multidimensional terms. Using descriptive statistics and social network analysis, I examine the teaching practices employed by a group of science and engineering faculty ($n = 56$). Results indicate the extensive use of lecturing with premade visuals (observed in 65% of all 2-minute intervals comprising a class). However, the majority of instructors ($n = 34$) lectured for periods of 20 minutes or less. Using the Differentiated Overt Learning Activities (Chi & Wylie, 2014) framework to interpret TDOP codes, the data reveal lower rates of active learning modalities including "being active" (students answering questions, 28%; students problem solving (PS),15%), "being constructive" (students asking questions, 4%; students doing creative tasks, 2%), and "being interactive" (students working with peers to do creative tasks, 2%). Results indicate variation across disciplines and course contexts, that active learning is embedded within PowerPoint lectures, and that small group work exercises are not synonymous with constructivist activities. Implications for research, practice, and policy are discussed. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99**:783–818, 2015

Correspondence to: Matthew T. Hora; e-mail: hora@wisc.edu

INTRODUCTION

The goal of this paper is to introduce the Teaching Dimensions Observation Protocol (TDOP) – and the instructional systems-of-practice framework upon which it is based – and describe how it can be used to document classroom dynamics in a nuanced and rigorous manner. This approach is based on the premise that educational practice is best viewed as "distributed in the interactive web of actors, artifacts, and the situation" as opposed to a phenomenon that can be understood through an individual's behavior regardless of the social and organizational context (Spillane, Halverson, & Diamond, 2001, p. 23). Such a perspective is important and necessary because classrooms are complex environments that involve not only instructor behaviors but also other actors and artifacts, which collectively interact over the course of a class period (Cohen & Ball, 1999). Yet commonly used surveys and observation protocols reduce teaching in ways that collapse time and the dynamism of teaching into coarsely-grained and unidimensional descriptors that largely focus on instructor behaviors alone. Such views of teaching have led to the widespread notion that teaching can be adequately described using terms such as "lecturing" or "active learning" (e.g., Freeman et al., 2014). Unfortunately, this approach leads to another, more problematic view of teaching: the conflation of coarse descriptors of teaching (e.g., lecturing) with particular modes of cognitive engagement. The instructional systems-of-practice framework offers a way to describe, in a more rigorous, fine-grained, and theoretically informed fashion, the complexity of teaching "in the wild" of real-world classrooms.

Why is a descriptive approach to the study of teaching that captures its multi-dimensional nature necessary? First, descriptive research is important to any arena of scientific inquiry to better understand the nature of the phenomenon being studied. For example, it almost goes without saying that biological scientists should describe the subjects of their research (e.g., organisms, animal behaviors, or entire ecosystems) in as precise a manner as possible. It is important that researchers studying problems in educational settings approach issues of research design and measurement the same care and precision that scientists utilize in their own disciplinary research (Derting, Williams, Momsen, & Henkel, 2011). Furthermore, given evidence that instructional activities that foster students' active engagement with the material and one another (hereafter referred to as a broad class of activities called "active learning") are more effective than passively listening to a lesson (Bransford, Brown, & Cocking, 1999), researchers are being encouraged to document the degree to which active learning modalities are being used, particularly in college and university classrooms (American Association for the Advancement of Science [AAAS], 2012). Yet currently available research instruments are not designed to describe and document nuances of classroom practices in general and subtle features of active learning in particular at a fine-grained level, particularly as they occur in naturalistic settings.

A second reason why descriptive research on classroom teaching is warranted pertains to the challenges associated with educational reform, particularly at the postsecondary level. While some signs indicate an increase in the adoption of active learning techniques (Hurtado, Eagan, Pryor, Whang, & Tran, 2012), other evidence suggests that instructors are not widely taking up these instructional practices (Henderson & Dancy, 2009; PCAST, 2012). Similar to educational reform efforts in K-12 settings, this state of affairs suggests that a gap exists between what researchers and policymakers consider "best practices" and what is actually being done in the classroom, or what is commonly viewed as the "fidelity" with which these practices are actually being used in the classroom. Yet coarsely-grained approaches to the study of teaching are not sufficient for assessing the "uptake" of active learning techniques because these rather complex approaches to curriculum and instruction focus not only on changing "surface structures and procedures" but also on altering

instructors' beliefs, norms of student-teacher interactions, and pedagogical principles embodied in the curriculum (Coburn, 2003, p. 4). Finally, in-depth accounts of teaching can be used as the basis upon which to design new programs or interventions so that they are responsive to local practices instead of top-down mandates which may alienate instructors (Cohen & Ball, 1999; Spillane et al., 2001).

Thus, the field of science education would benefit from a way to describe and document classroom teaching in a way that maintains fidelity to its complex and dynamic nature, while also being able to discern the presence (or absence) of active learning modalities. In response, the TDOP was created as part of a study that explored the cognitive, cultural, and contextual characteristics shaping instruction in postsecondary institutions.[1] The TDOP captures key elements of instructional systems-of-practice theory through its focus on fine-grained descriptors of classroom dynamics, interactions among actors and artifacts, and the temporal fluctuation of these phenomena over time. Since its initial development in 2008, the instrument has been widely adopted with users particularly appreciating the detailed and nonevaluative nature of the TDOP based in large part on the view that postsecondary faculty[2] are especially resistant to having the "quality" of their teaching be determined by a single score or rubric (Chism, 2007) and also the perceived need in the field for more rigorous methods to measure teaching as an empirical phenomena (AAAS, 2012). The TDOP has been featured in several research papers (e.g., Clark, Norman, & Besterfield-Sacre, 2014; Code, Piccolo, Kohler, & MacLean, 2014; Finelli, Daly, & Richardson, 2014), adapted by other researchers (Smith, Jones, Gilbert, & Wieman, 2013)[3] used in numerous studies focused on science education, and over 300 researchers have used the online version of the instrument (http://tdop.wceruw.org).

While research conducted with the TDOP has shed light on the insufficiency of descriptors such as "lecturing" (Hora & Ferrare, 2014) and documented in teaching across disciplinary groups (Hora & Ferrare, 2013), one issue has been underexamined—whether the instructional systems-of-practice framework and the TDOP can contribute insights to the prevalence and nature of active learning. In this paper, I describe techniques for combining TDOP codes to detect the presence of active learning using the Differentiated Overt Learning Activities (DOLA) framework, which is a taxonomy of classroom-based active learning modalities based on observable student behaviors (Chi & Wylie, 2014). Using the DOLA framework to organize combinations of TDOP codes, I analyzed data from observations of 56 instructors using descriptive statistics and social network analysis to answer two research questions: (1) What teaching practices are employed by a group of science and engineering faculty? (2) What is the prevalence and nature of active learning observed in these classrooms? In addition, I also analyze the data within different class sizes and course levels based on the role that the organizational context plays in shaping practice as theorized by the instructional systems-of-practice approach.

[1]The development of the TDOP has been a group endeavor, with Joseph J. Ferrare acting as a co-developer from 2008 to 2013. In addition, a team of four researchers including myself, Jana Bouwma-Gearhart, Amanda Oleson, and Jennifer Collins collected data in the spring of 2013 that are reported in this paper. While this paper is single authored the study reported here would not be possible without the contributions of these valued colleagues.

[2]By faculty I mean all people who hold undergraduate teaching positions—whether full- or part-time, tenured, or untenured—in postsecondary institutions with the exception of graduate teaching assistants.

[3]This effort led to a protocol known as the COPUS, which is a minor adaptation of the TDOP that involved removing certain categories (e.g., student cognitive engagement) and re-naming existing codes.

BACKGROUND

In recent years, the postsecondary science classroom has been the focus of extensive educational reforms promoting instructional activities that more directly engage students in their own learning. Indeed, the federal government views the adoption of active learning in postsecondary science, technology, engineering, and mathematics (STEM) classrooms as a critical national priority associated with economic development, diversity in the workforce, and the public's scientific literacy (e.g., PCAST, 2012). Much of this focus is a reaction to the persistence of didactic lecturing in higher education, where faculty act as a "sage on the stage" by transmitting knowledge to their passive students (Mazur, 2009). This reaction has been fueled by growing evidence in cognitive psychology and the learning sciences (Bransford et al., 1999; Chi & Wylie, 2014), and discipline-based education research (National Research Council, 2012) that instructional activities that directly engage students in actively constructing their own understanding of course material are more effective than passive modes of learning (e.g., Duch, Groh, & Allen, 2001; Crouch & Mazur, 2001).

Given that funding agencies and educational leaders are investing substantial resources to support the nationwide adoption of active learning techniques, these groups have a growing interest in documenting the extent to which the resources are being used in the nation's colleges and universities (AAAS, 2012). This raises a key question: what do we really know about how science is taught in undergraduate classrooms? In the most comprehensive analysis of faculty teaching in the United States, the Higher Education Research Institute (HERI) conducts a survey of faculty work that investigates teaching methods, advising practices, and job satisfaction. In the 2010–2011 survey, which included 23,824 full time undergraduate faculty at 417 postsecondary institutions, researchers found that STEM faculty ($n = 6768$) reporting using "extensive lecturing" in 63% of the courses they taught, "class discussion" in 61.5%, "cooperative learning (small groups) in 47%, and "using student inquiry to drive learning" in 36.5% of their courses (Hurtado et al., 2012).[4] Unfortunately, what terms such as "extensive lecturing" means in operationally precise terms is not clear in the survey. A smaller survey-based study focusing on physics instructors found that among 722 instructors, 29% reported using peer instruction, 13.9% reported using interactive lectures, and 13.7% reported using cooperative group problem solving (PS) (Henderson & Dancy, 2009).

While evidence from self-reported surveys such as these shed light on the nature of teaching in STEM classrooms, they are limited in several ways. In a critical review of surveys used to study college student experiences, Porter (2011) highlighted several validity-related issues such as the ability of respondents to accurately and reliably recall their experiences over imprecise periods of time. This concern about the process of respondent recall and the granularity of resulting data is salient for surveys such as the HERI Faculty Survey that asks faculty to report their teaching across "most classes." The widespread reliance on self-reported data to document teaching is also limited by the lack of research exploring the congruence (or lack thereof) between self-reports and actual classroom practice (Kane, Sandretto, & Heath, 2002; Mayer, 1999). Thus, questions remain about precisely what is meant by data such as "63% of faculty are extensive lecturers" in terms of specific classroom behaviors and whether the claim actually reflects respondents' teaching practices.

[4]When the final version of this manuscript was being prepared, data from the 2013–2014 HERI Faculty Survey had not yet been released that disaggregated results by disciplinary group. Thus, in this paper I report data from the 2010–2011 version of the survey.

Problems With Conceptualizing and Operationally Defining Classroom Teaching

But perhaps the biggest issue facing the field pertains to how teaching itself is conceptualized and operationalized in research instruments.

Reliance on Ill-Defined Descriptors of Teaching

One of the challenges in studying teaching is researchers' tendency to characterize instruction solely in terms of teaching methods such as "lecturing" or "small group work." This tendency is problematic because such descriptions frequently lack accompanying definitions that specify which types of practices are encompassed within terms such as a "lecture." An operationally precise definition of lecturing, for example, would indicate the length of time an instructor must be speaking to be considered a lecture, whether a period of verbal exposition interspersed with questions or other activities constitutes a lecture, and so on (Schonwetter, 1993). This precision is important because within a lecture an instructor may actually be utilizing principles of effective instruction or setting the stage for interactive modalities (Perry & Smart, 1997; Saroyan & Snell, 1997; Schwartz & Bransford, 1999). Indeed, in practice some active learning techniques such as Scientific Teaching do not entirely preclude the use of lecturing but instead aim for only 34% of the class period to be devoted to lecture, ideally broken into segments no longer than 10 minutes (Miller, Pfund, Pribbenow, & Handelsman, 2008). Guided inquiry is an approach where instructors conduct mini-lectures as a way of providing a scaffold to more constructivist activities (Hmelo Silver, Duncan, & Chinn, 2007). Finally, Crouch and Mazur (2001, p. 975) note that in Peer Instruction classrooms, lecturing also takes place: "We typically devote one-third to one-half of class time to ConcepTests and spend the remainder lecturing." In these cases the lecture serves as a way to introduce new topics and to prepare students for more in-depth activities, which highlights the fact that "lecturing" can serve a pedagogically useful purpose.

Ultimately, the lack of carefully specified definitions for teaching methods is an issue because it results in the absence of a shared view of what these methods mean in practice among the research community (Menekse, Stump, Krause, & Chi, 2013). As a result, researchers often fail to articulate what specific behaviors constitute lecturing and other types of teaching. For example, in an experiment comparing lecturing to interactive teaching, the lecturing condition is defined as "using PowerPoint slides to present content and example problems and also (showing) demonstrations" (Deslauriers, Schelew, & Wieman, 2011, p. 862), a definition which encompasses a diverse array range of instructional behaviors and technologies. Perhaps more problematic is the definition provided for the experimental condition (i.e., interactive teaching): "There was no formal lecturing; however, guidance and explanations were provided by the instructor throughout the class" (Deslauriers et al., 2011, p. 863). The exact difference between instances where the instructor provides "guidance and explanations" and is engaged in "formal lecturing" is not provided. Such an approach, in failing to articulate the precise nature of the experimental and control conditions, ultimately raises questions about the validity of the results (Derting et al., 2011; Hora, 2014a), which may be one reason for discrepancies across studies comparing different modes of teaching (e.g., Kirschner, Sweller, & Clark, 2006).

Researchers also commonly describe teaching in binary terms, with mutually exclusive categories such as "active" or "passive" teaching (Menekse et al., 2013) or "lecturing" and "active learning" (Freeman et al., 2014). Such an approach echoes findings from research on faculty cognition that posits instructors' thinking about teaching can be described as either

student centered or instructor centered (Kember, 1997). However, this body of research has been critiqued for its lack of ecological validity, as "a strong opposite 'either/or' positioning of the approaches does not do justice to the nature of the phenomenon" (Postareff & Lindblom-Ylanne, 2008, p. 120). Instead, in practice faculty actually exhibit multiple, often contradictory beliefs about learning that are closely linked to the task at hand (Hora, 2014b). In terms of teaching, this more nuanced view suggests that while some instructors may solely use a "pure" lecturing or active learning approach, there are more subtle variations of these categories in real-world classrooms in terms of time allocated to different types of teaching as well as underlying pedagogical intentions.

To understand the nature of classroom teaching and learning at a finer grain size, Chi (2009) developed the DOLA framework to deconstruct the descriptor "active learning." Indeed, the different active learning traditions noted above (e.g., Scientific Teaching, Peer Instruction) involve distinct approaches to course design and classroom teaching. The DOLA framework was based on the premise that descriptors such as active learning limit the field in a number of ways, including the failure to specify conditions for experimental research in precise terms, and by providing educators with overly ambiguous language to use when attempting to change their teaching from passive to active modalities (Chi & Wylie, 2014). A core feature of the framework is the ICAP hypothesis, which posits that three distinct types of student engagement in the classroom exist in ascending order of efficacy: being active (where students are visibly engaged in activities that activate their own knowledge); being constructive (where students are visibly engaged in activities where they generate their own knowledge); and being interactive (where two or more students are visibly engaged in activities that develop knowledge). The evidentiary base indicates that each of these modalities is more effective than when students are "being passive" in the classroom, but that important differences exist among them as well (Chi, 2009; Chi & Wylie, 2014). This framework is predicated on the notion that while challenging for researchers, these different modalities can be empirically observed in the classroom by documenting the overt behaviors of students. As a result, the DOLA framework represents a significant advance toward the goal of providing researchers with the tools necessary to detect subtle features of active learning in the classroom.

Assuming Student Cognition From Ill-Defined Descriptors of Teaching

From this reliance on reductionist and binary descriptors of teaching arises the second issue: the conflation of specific instructional activities with distinct modes of student cognition (Chi & Wylie, 2014). That is, it is not uncommon to see researchers claim that a particular instructional practice (e.g., small group work) is synonymous with a specific mode of student cognitive engagement such as actively constructing new knowledge. As Chi and Wylie (2014, p. 235) observe, "simply asking students to work together does not automatically make an activity interactive." One of the reasons such an assumption is mistaken is the inherent variability in how different instructors may implement a particular instructional technique. For example, in a study on how the Peer Instruction method (Crouch & Mazur, 2001) was being implemented in undergraduate physics classrooms, Turpen and Finkelstein (2009) found substantial variation in the ways in which questions were asked and in the subsequent learning opportunities afforded to students (see also Zhang & Linn, 2013).

Beyond considerations of instructors' various ways of teaching, however, is the long-standing empirical problem of discerning what is going on in students' minds during a lesson (Nystrand & Gamoran, 1991). Thus, the field of science education are faced with the problem of how to discern the relationship between types of student cognitive engagement

and instructional activities without falling prey to the trap of assuming that one always and unequivocally predicts the other. While exploring these dynamics in controlled, lab-based settings is important, so too are studies that examine the relationship between teaching and cognition "in the wild" of actual classrooms. Such research can serve the pressing need to describe and document instructional practice in the nation's postsecondary classrooms (PCAST, 2012) while accounting for the role of local contexts in shaping how students experience different types of teaching. To achieve such fine-grained descriptions, one method is particularly well suited to this task: classroom observations.

Using Classroom Observation Protocols to Study Teaching

Classroom observation is a technique for collecting educational data where researchers take notes and/or code teacher and student behaviors in actual classrooms or from videotaped lessons. As such, classroom observations are part of a larger class of research tools for studying behavior that is used in fields such as cultural anthropology, consumer behavior, and ethology. Classroom observations are used across the educational spectrum from elementary to postsecondary institutions, but the technique is more pervasive and instruments more rigorously tested in K-12 settings, where protocols such as the CLASS protocol (Pianta & Hamre, 2009) and Charlotte Danielson's Framework for Teaching (Danielson, 2013) are central features of teacher evaluation systems in districts across the country. Observations in postsecondary settings are generally used in less high-stakes situations such as professional development and peer mentoring (Chism, 2007). For these applications, the protocols used are often unstructured rubrics where observers take notes but have no prespecified directions about what behaviors or facets of teaching to record and in what fashion.

As interest in the quality and efficacy of postsecondary instruction has increased in recent years, more structured observation protocols have been introduced to the field. The Teaching Behaviors Inventory (TBI) was one of the first widely used protocols in postsecondary settings (e.g., clarity and organization). The TBI is a 60-item instrument composed of eight categories of teaching that requires observers to assign a score on a five-point scale ranging from "almost never observed" to "almost always observed" at the conclusion of the class (Murray, 1983). Another instrument is the Reformed Teaching Observation Protocol (RTOP), which has its basis in the standards-based reform movements in science and math education (MacIssac & Falconer, 2002). The RTOP consists of 25 items scored at the end of a class, which can then be used to classify instructors into one of five categories. Two of these categories represent teacher-centered classrooms (e.g., category one represents "straight lecturing") and three represent learner-centered classrooms (e.g., category five represents "active student involvement in open-ended inquiry") (Ebert-May et al., 2011). Researchers have also developed several other protocols that are similar in structure and intent (e.g., Walkington et al., 2011).

Three characteristics of protocols such as the TBI and RTOP are worth noting. First, in assigning single scores at the end of a class period, these instruments ignore the role of time and the duration with which specific teaching behaviors are observed. Second, while these instruments are designed to capture distinct dimensions of teaching, they capture relatively coarse measures of instruction. Further, the summative scoring procedure makes it impossible to explore the interactions among these dimensions over time. Finally, it is important to note that ultimate purpose of protocols such as the RTOP is to evaluate the quality of instruction rather than to describe it, which is an approach not without limitations.

A recent review of the reliability of evaluative protocols found that ratings of individual teachers varied considerably across analysts (Guarino & Tracy, 2012), and that rater bias (e.g., preexisting beliefs about what constitutes high-quality teaching) is a major reason for

the high degree of variation (Cash, Hamre, Pianta, & Meyers, 2012). Partly in response to these concerns, as well as to the perceived need for more rigorous descriptions of practice, researchers have recently developed a class of observation protocols that focuses on describing teaching in fine-grained terms, including the TDOP and the Real-Time Instructor Observing Tool (West, Paul, Webb, & Potter, 2013). These newer, more descriptive approaches are also motivated by a desire to shift the analytic focus from the instructor alone to a more comprehensive and systemic account of the classroom.

The Teaching Dimensions Observation Protocol

The development of the TDOP was largely inspired by a growing movement in educational research that focuses on describing practice as it unfolds in real-world settings, rather than attempting to disseminate best practices with little attention to local contexts (Coburn, 2003; Spillane et al., 2001). Descriptive research of educational practice has been largely motivated by the realization that the implementation of policy as well as the adoption (or rejection) of curricula or teaching methods is strongly shaped by the cultural norms, routines, and structural constraints within particular school settings (Spillane, Reiser, & Reimer, 2002). Practice-oriented scholars have drawn on theories of situated and distributed cognition that emphasize the interdependence of individuals' cognitive processes and the environment in shaping how individuals make decisions, the nature of knowledge, and learning itself (e.g., Brown, Collins, & Duguid, 1989) to study educational practices such as principals' administration of teacher evaluation policies (Halverson & Clifford, 2006), individuals use of mathematics in real-world settings (Lave, 1988), and teachers' decisions about in-class teaching (Schoenfeld, 1999).

In much of this work scholars pay close attention to the role that intentionally designed tools or artifacts (e.g., technology) play in mediating activity (Wertsch, 1991). This mediation process occurs through users' perceptions of limited avenues of potential behaviors (Greeno, 1998) and by artifacts acting as "scaffolds" for learners to perform tasks beyond their existing capacities (Pea, 1993). For example, an individual may perceive that a chair is for sitting, or a teacher may perceive that PowerPoint slides afford the distillation of knowledge into bullet points (Adams, 2006). Ultimately, this perspective emphasizes the interdependence of individuals and their environment in shaping *why* people act the way they do as well as *what* constitutes practice itself.

In a study on how leaders develop professional communities within schools, Halverson (2003, p. 3) built on these ideas by advancing the systems-of-practice approach, which emphasized how networks of tasks, structures, and artifacts (e.g., tools, policies, and procedures) created "complex webs of practice in organizations." Besides the critical role that artifacts play in shaping practice, the systems-of-practice approach also emphasizes the role of time at both the microlevel in terms of task performance and the macrolevel in terms of cultural norms that develop over time as activities are repeated within particular groups and settings. Given its focus on describing educational practice in such a comprehensive manner, my colleagues and I adapted Halverson's (2003) framework and applied it to the study of classroom teaching.

This process began with an instrument originally designed to study inquiry-based science teaching in middle schools that reflected core ideas of the systems-of-practice approach, primarily due to its inclusion of multiple categories to characterize instruction and the innovative use of a time-sampling framework (Osthoff, Clune, Ferrare, Kretchmar, & White, 2008). Using this protocol as a foundation for the TDOP, we identified five key aspects of classroom dynamics: *teaching methods* (e.g., small-group discussion), *pedagogical strategies* used in the classroom (e.g., organization), types of *student–instructor interactions*

in the classroom (e.g., types of questions posed), the types of *cognitive engagement* that instructors place on students, and the use of *instructional technology* (e.g., chalkboards). Each category contains several codes that represent specific, overt, and observable behaviors of the instructor and their students in the classroom. It is important to note that the categories do not represent latent constructs but instead are simply groups of codes that capture distinct aspects of teaching (see Table 1).

Two of the categories featured in the TDOP require further examination. First, the teaching methods category includes several codes that encompass instructors' engagement in verbal exposition, which is commonly known as lecturing. In the TDOP, this mode of instruction is decomposed into distinct types of lecturing that implicate other people (e.g., Socractic lecturing) or specific tools and artifacts (e.g., premade visuals such as PowerPoint slides). The latter aspect of lecturing is particularly important given the focus on artifact use as a key mediator of activity in the instructional systems-of-practice framework. Second, the cognitive engagement category refers to the types of cognitive activity that students may be experiencing in the classroom. This category is based on research demonstrating that the type and degree of student cognitive engagement in the classroom is a key aspect of learning (Blumenfeld, Kempler, & Krajcik, 2006; Chi & Wylie, 2014). Measuring cognitive engagement is inherently difficult, and measurement strategies include inferring student engagement from overt student–instructor interactions or in-class learning activities (Nystrand & Gamoran, 1991). Despite the challenges associated with inferring student cognition, it is an important dimension of instruction to capture and is one of the distinguishing features of the TDOP in comparison to other descriptive observation protocols that use a similar time-sampling framework (Smith et al., 2013; West et al., 2013).

While the study reported in this paper is not a validity study for the TDOP, it is important to address issues related to validity and reliability for new research instruments. Traditionally, validity in higher education has focused on establishing criterion validity (how well a score predicts or estimates a measure that is external to the test) and construct validity (how well a measure adequately captures the domain of interest) for surveys. Increasingly, scholars have adopted argument-based approaches to validity, which entails collecting varied sources of evidence and theory to support the interpretation of particular measures in light of their intended uses (Kane, 2001; Porter, 2011). Given the intended use of the TDOP to provide descriptive accounts of teaching and not to ascertain the presence of an external criterion, testing for criterion validity was not appropriate. Instead, we tested face and construct validity for each of the codes and categories through preliminary fieldwork and feedback from disciplinary and education experts. These groups of faculty confirmed that the codes included in the instrument were consistent with their own understanding of teaching. Additionally, since groups of codes are not intended to measure latent constructs, construct validity tests on this point were not applicable. Perhaps most importantly for observation instruments being used by multiple raters is interrater reliability (IRR), which ensures that different analysts will use an instrument in a similar manner across cases. The training procedure for the TDOP is thus rather extensive and places considerable focus on developing IRR. As further development with the TDOP continues, additional validity and reliability evidence will need to be gathered (e.g., test–retest reliability).

METHODS

This study took place at three large, public research universities in the United States and Canada in the spring of 2013. The three study sites had similar undergraduate enrollments and external research funding. We selected these research universities in part because of

TABLE 1
Description of TDOP Categories and Codes

TDOP Category	Code	Description of Code
Teaching methods		
Lecturing	L	Instructor speaks to students with no media
Lecturing with premade visuals	LPV	Instructor speaks with premade visual media (e.g., PowerPoint slides)
Lecturing with hand-made visuals	LHV	Instructor speaks to students with hand-made visuals (e.g., writing on chalkboard)
Lecturing with demonstration	LDEM	Instructor speaks while using demonstrations
Socratic lecture	SOC-L	Instructor speaks while asking questions (two or more), the answers to which guide the discussion
Working through problems	WP	Instructor works out computations or problems
Small group work	SGW	Students form into groups of 2^+
Desk work	DW	Students complete work alone at desk
Multimedia	MM	Instructor plays a video/movie without speaking
Assessment	A	Instructor gathers student learning data
Pedagogical moves		
Humor	HUM	Instructor tells jokes (2^+ students must laugh)
Anecdote/example	ANEX	Examples that link material to student experiences
Graphic	GR	Instructor uses graphic image to illustrate material
Organization	ORG	Instructor clearly indicates transition between topics
Emphasis	EMP	Instructor clearly states a topic is important
Instructor–student interactions		
Rhetorical questions	IRQ	Instructor poses questions without waiting for answer
Display questions	IDQ	Instructor poses questions seeking information
Comprehension questions	ICQ	Instructor poses question about student understanding
Student novel question	SNQ	Student asks original question
Student comprehension question	SCQ	Student asks for clarification about previous topic
Student response	SR	Student responds to instructor question
Student peer interactions	PI	Students interact with one another
Cognitive engagement		
Problem solving	PS	Students are asked to actively solve a closed-ended problem with a known solution
Creating	CR	Students are asked to actively solve an open-ended problem without a known solution

(*Continued*)

TABLE 1
Continued

TDOP Category	Code	Description of Code
Connecting to real world	CN	Students are given examples linking material to common experiences
Instructional technology		
Chalkboard	CB	Chalkboard or whiteboard used for writing
Overhead projector	OP	Machine used to project images on screen
PowerPoint	PP	Microsoft PowerPoint slides
Clicker response systems	CL	Clicker response systems
Demonstrations	D	Laboratory demonstration equipment
Digital tablet/document camera	DT	Machine used to project images and writing on screen
Movies	M	Movies (e.g., YouTube movies)
Simulations	SI	Graphic simulations and animations

the large number of undergraduates being trained in STEM disciplines at these institutions, as well as the fact that all three had STEM pedagogical improvement initiatives underway at the time of data collection. Personnel active in these initiatives provided initial contacts for our team of researchers.

Faculty were included in the study population if they were listed as instructors in the course schedule for the spring semester. One hundred sixty-five individuals were contacted via email with a request to participate in the study, and 56 ultimately participated (34% response rate). The participants represented the following disciplinary groups: biology ($n = 18$), mechanical engineering ($n = 12$), geology ($n = 15$), and physics ($n = 11$). We selected these disciplines due to the large populations of instructors across the study sites and for their leadership in STEM education initiatives.[5] Faculty self-selected into the study, and thus the results should not be generalized to the larger population of instructors at these institutions or in higher education (see Table 2). It is important to note that the percentage of instructors not on the tenure-track represented in this study (38%) was roughly in line with the proportion of contingent faculty at participating institutions where data were available (i.e., 33% and 47%).

The course component of interest in this study was the class period, colloquially known as the "lecture" period. That is, laboratory and discussion sections were not observed. Thirty-four lower division and 22 upper division courses were included in the study, the designation of which was determined by consulting each institution's course numbering system (e.g., lower division courses at one institution were numbered 1000–2000 and upper division courses 3000–4000). The courses also varied by enrollment numbers, which were obtained from the instructor.

Training Procedures for the TDOP

A team of four researchers collected data at the study sites during weeklong field visits in the spring of 2013. Prior to gathering data, the four researchers participated in a rigorous training program that took approximately 28 hours of training spread out over 2 weeks

[5]Given that these disciplines reflect only a few of fields captured within the acronym of "STEM," for the remainder of the paper I refer to science and engineering disciplines.

TABLE 2
Description of Sample

	Participant (*n*)	Percentage
Total	56	100
Sex		
Female	18	32
Male	38	68
Discipline		
Biology	18	32
Mechanical engineering	12	21
Geoscience	15	27
Physics	11	20
Level of course		
Lower division	34	61
Upper division	22	39
Size of course		
25 or less	8	14
26—100	18	32
101—199	18	32
200 or more	12	22
Position type		
Lecturer/instructor	21	38
Assistant professor	9	16
Associate professor	14	25
Professor	12	21

to reach acceptable IRR.[6] The first phase of training in the TDOP involved introducing the raters to each of the codes and rules for applying them in different situations, as well as the process for applying the codes in the classroom. This introductory meeting was followed by two 3-hour long sessions where the entire group coded 10–15 minute segments of videotaped classes, followed by extensive discussions about areas of disagreement while the group also developed a shared understanding of coding rules and procedures. All videotapes used for the training were downloaded from the YouTube channels of research universities. Videotapes for the training were also selected from large physics and biology courses (i.e., "lecture" classes in stadium-style classrooms) to account for variability across disciplines.

Next, the group coded two 50-minute long classes that were followed by formal IRR testing using Cohen's kappa. Cohen's kappa is an index that measures the level of agreement between two sets of dichotomous ratings, while taking into account the possibility that agreement can take place by chance. The calculations for kappa were made between all possible pairs of raters (i.e., six pairs), with kappa values disaggregated at the code

[6]It should be noted that the claim made in the Smith et al. (2013) paper that adequate training to establish IRR using an abridged version of the TDOP can be conducted in 1.5 hours, runs counter to the experiences of the research group whose work is reported in this paper as well as others experienced in conducting reliability trainings for observation-based research (Joe et al, 2013). While the extensive nature of the training for the TDOP reported in this paper may appear daunting, it is important to recognize that collecting behavioral data in a scientifically rigorous manner is no easy task. As such, drastic reductions in training time made to simplify data collection procedures for researchers should be carefully weighed against potential costs in terms of the quality and utility of the data. That said, our group is actively developing an online training module that will reduce the training time described in this paper.

TABLE 3
Description of TDOP Interrater Reliability Scores for Analysts

Analyst	Teaching Methods	Pedagogical Moves	Interactions	Cognitive Engagement	Instructional Technology
Analyst 1/analyst 2	.90	.85	.83	.74	.94
Analyst 1/analyst 3	.82	.81	.73	.78	.90
Analyst 1/analyst 4	.89	.74	.79	.71	.90
Analyst 2/analyst 3	.83	.80	.81	.75	.89
Analyst 2/analyst 4	.84	.75	.79	.77	.89
Analyst 3/analyst 4	.80	.73	.72	.74	.91

category level (e.g., teaching methods) to make it possible to assess raters' agreement on each dimension. It is important to note that the data structure for calculating kappa scores is laid out as a table such that rows represent each code at a particular 2-minute interval and each column is one of two raters. Thus, each row is an "event" indicating whether each rater coded that code as being present in the interval or not. These calculations were all conducted using the TDOP Web site, which has a built-in capability for testing IRR. This process of coding two 50-minute long classes, calculating kappa scores across all rater pairs, examining areas of disagreement, and then returning to another round of coding two new classes took place three times until adequate kappa values were obtained (see Table 3).

Through this intensive training process and formal IRR testing, the reliability of the evidence gathered for the study was demonstrated. In future applications of the TDOP, the training procedures described above are being enhanced by following training protocols used in large-scale observation research. These procedures include testing reliability using a larger number of testing videos (i.e., 5% of total dataset) and testing each rater against precoded videos that are considered to be the "standard," representing perfect application of the instrument as determined by a group of experts (e.g., Joe, Tocci, Holtzman, & Williams, 2013).

Data Collection

After obtaining permission from the instructor, analysts sat near the back of the classroom and then proceeded to observe the entire class period. A total of 95 class periods were observed, with 39 faculty observed twice and 17 faculty observed once. This discrepancy was due to scheduling issues such as exam dates and courses that met only one time a week. The study team used the online form of the TDOP that entails clicking on a code when it is observed during a given 2-minute interval. It is important to note that because a variety of practices may occur within a single interval, more than one code for a given dimension may be coded within the same interval. Furthermore, in instances where a behavior started in one interval (e.g., 2:00–3:59) and ended in another (e.g., 4:00–5:59), it was coded in both intervals. While this coding procedure may seem overwhelming, with adequate training the coding scheme and corresponding demands on the observer is not an issue. That being said, the challenges inherent in using a time-sampling protocol underscore the importance of rigorous training. Finally, for the training and the fieldwork the team used a form of the TDOP that included 47 codes, but in this paper data are reported for only 33 of these codes. The 14 removed codes were primarily instructional technology codes that were not regularly observed, and their removal allows for a more concise presentation of study results.

Data Analysis

Data from the TDOP instrument were exported from the Web site into spreadsheets where individual 2-minute intervals are rows and codes are columns. A code observed by the analyst is indicated by a "1" and not observed is represented by a "0." These raw data were analyzed by calculating simple proportions of individual codes, predetermined code combinations (i.e., TDOP codes mapped onto the DOLA framework), and social network analyses of data from different course levels (i.e., lower and upper division).

Identifying Frequencies for Individual Codes

First, I report the proportion of times that a particular code was observed across all 2-minute intervals. All figures are rounded to two decimal places. These data are reported for the entire sample as well as by groups that reflect potentially important points of variability within academic contexts including discipline, course level, and class size. It is important to note that analysts scored codes as present if the corresponding practice was observed for any portion of a given 2-minute interval. Thus, the frequencies reported reflect the portion of intervals in which the code was observed, but only roughly approximate the amount of actual class time in which the code occurred. Furthermore, since multiple codes can occur simultaneously, the sum of the various interval codes typically exceeds the total amount of class time. In addition to reporting simple proportions, I calculated the extent of "straight lecturing" in the study sample, which is defined as periods of verbal exposition (as captured by any of the lecturing codes) where none of the codes signifying active learning are present (e.g., student response [SR], problem solving (PS), students' novel questions [SNQ], student response (SR), creating (CR), and student peer interactions (PI)). The raw TDOP data were analyzed to categorize instructors based on the maximum length of time they spent teaching in this manner (i.e., up to 20 minutes, 21–40 minutes, over 40 minutes).

Identifying the Prevalence of Active Learning

Next, to identify the prevalence of overt student behaviors, indicative of different types of active learning, codes from the TDOP were mapped onto the three types of active learning discussed in the DOLA framework (Chi & Wylie, 2014). Originally, an attempt was made to distill aspects of well-known active learning strategies such as peer instruction (Crouch & Mazur, 2001) and scientific teaching (Handelsman et al., 2004) that could be linked to TDOP codes, but this approach was rejected because these strategies were not easily operationalized in terms of overt behaviors such that the underlying intent and breadth of the strategy could be captured. Other strategies, such as collapsing descriptive codes into global indicators of student-based activity and instructor-based activity (e.g., Smith et al., 2013), were also not used due to the aforementioned limitations with reducing classroom complexity to such coarse categories as well as the limitations in assuming particular modes of student cognition based on the primary interlocutor in class.

While the TDOP was not developed with either the DOLA framework or other categories of active learning in mind, several TDOP codes focused on student behaviors were well suited to the three categories identified by Chi (2009). The mapping of the two approaches was also made possible by the focus on visible engagement in specific activities emphasized in the DOLA framework. As Chi and Wylie (2014, p. 220) note, while "far from perfect, overt behaviors are a good proxy to reflect different modes of engagement." The following cross-protocol correspondences were then identified (see Table 4).

TABLE 4
Description of TDOP Code Combinations Used to Capture Elements of the Differentiated Overt Learning Activities (DOLA) Framework (Chi & Wylie, 2014)

DOLA Category	TDOP Category and Code
Being active	Student–instructor interactions: Student response (SR) Or Cognitive engagement: Problem solving (PS)
Being constructive	Student–instructor interactions: Student novel questions (SNQ) Or Cognitive engagement: Creating (CR)
Being interactive	Student–instructor interactions: Student peer interactions (PI) And Cognitive engagement: Creating (CR)

Being Active. The "being active" modality is defined as students being visibly engaged in activities that activate their own knowledge related to course content. Examples from the DOLA framework include following experimental procedures, repeating ideas out loud, highlighting while reading, and copying solutions from chalkboard. To capture the *being active* mode, two TDOP codes were identified (SR or PS). The codes included in this category capture question-and-answer exchanges between students and teachers, with a focus on whether the student responds to an instructor's question. Then, the cognitive engagement mode of problem solving indicates that students are being observed actively engaged in working on closed-ended problems, often (but not always) via desk work (DW).

Being Constructive. The "constructive" modality is defined as students visibly engaged in activities where they generate their own knowledge beyond materials presented in class. Examples from the DOLA framework include self-explaining, drawing concept maps, asking comprehension questions, solving problems requiring the construction of knowledge, and designing a study. To capture the *being constructive* mode, two TDOP codes were identified (SNQ or CR). SNQ refers to instances where students were observed asking original questions to the instructor, as opposed to simply asking for clarification on a topic. The cognitive engagement code of CR refers to instances where instructors give students an open-ended task where no fixed answer exists.

Being Interactive. The "interactive" modality is defined as two or more students visibly engaged activities that develop understanding beyond materials presented in class. Examples from the DOLA framework include working in groups or pairs or interacting with feedback from the instructor. To capture the interactive mode, two TDOP codes were identified that must be observed together within a 2-minute interval (PI and CR) to be included in the *being interactive* category. PI refers to observable interactions between two or more students. Then, because it is possible for PI to occur without any sort of constructive aspect to the work and/or students engagement, the interactive modality is only incurred when PI is observed with the CR mode of cognitive engagement.

It is important to note that the TDOP codes selected for this exercise do not capture the breadth of active learning as posited by the DOLA framework. Future research and development will involve the identification of other overt behaviors that

can reliably capture additional aspects of active learning (e.g., student self-explanation, instructor use of varied representations). Also, these codes are not intended as latent constructs that indicate the presence (or absence) of instructional quality, but instead are best viewed as indicators of the three types of active learning as suggested by the DOLA framework. To identify the frequency with which these indicators of active learning were observed, I created one new variable using the "and" operator in SPSS statistical analysis software to capture frequencies for the PI and CR code combination.

Social Network Analysis

Next, I used techniques from social network analysis to delineate configurations within and between the dimensions of practice. The raw data for these analyses are in the form of two-mode (or "affiliation") matrices consisting of instructors' 2-minute intervals as rows (mode 1) and TDOP codes as columns (mode 2). Using UCINET (Borgatti, Everett, & Freeman, 2002), the two-mode matrix was transformed into a one-mode (code-by-code) matrix through matrix multiplication. This transformation results in a valued co-occurrence matrix in which each cell corresponds to the number of intervals in which two given TDOP codes are affiliated. For example, the intersection of the codes for SGW and PS could have a value of three, which means these two codes were cocoded in three intervals across all instructors in the matrix. I then used the program Netdraw to graph the co-occurrences between each pair of codes across all instructor intervals. The results from the valued co-occurrence matrix were also used to report the strength of ties between active learning indicator codes and other codes.

RESULTS

What Teaching practices are Employed by a Group of Science and Engineering Faculty?

To answer the research question pertaining to the types of teaching practices observed in the field, I first report results for each individual TDOP code for the entire sample and then by discipline, course level, and class size. Then, to illustrate the interconnected and temporal nature of distinct dimensions of instructional practice as suggested by systems-of-practice theory, I present a pair of graphs created using social network analysis techniques.

Frequency of Individual Codes

The data in Table 5 represent the proportion of times that each TDOP code was observed across all 2-minute intervals for the entire sample as well as for each disciplinary group.

These data highlight the prevalence of certain instructional practices across the five TDOP dimensions. Notable results include an extensive amount of lecturing with premade visuals (LPV, 64% of all 2-minute intervals), lecturing with hand-made visuals (LHV, 27%), the administration of assessments (A, 11%), and small group work (SGW, 11%). These results highlight the prevalence of instructional technologies used in the lecturing mode of instruction, particularly that of PowerPoint slides and chalkboards. Then, given the predominance of the "lecturing" descriptor to connote extensive periods of verbal exposition with no student engagement (i.e., a "straight" lecture), and the related goal for faculty to reduce the length such periods of exposition (e.g., Handelsman, Miller, & Pfund, 2007), the

TABLE 5
Classroom Observation Data Using the TDOP by Discipline

Discipline		All	Biology	Mechanical Engineering	Geo-science	Physics
Instructors		56	18	12	15	11
Total 2-minute intervals		2,514	751	527	767	469
Teaching methods						
Lecturing	L	.06	.02	.06	.10	.08
Lecturing with premade visual	LPV	.64	.86	.56	.63	.43
Lecturing with handmade visual	LHV	.27	.11	.52	.14	.49
Lecturing with demonstration	LDEM	.03	.02	.07	.01	.05
Socratic lecture[a]	SOC-L	.03	.06	.04	.02	.01
Working through problems	WP	.08	.02	.26	.00	.19
Small group work	SGW	.11	.11	.10	.13	.11
Desk work	DW	.07	.03	.12	.03	.12
Multimedia	MM	.03	.03	.04	.02	.02
Assessment	A	.11	.17	.08	.07	.13
Pedagogical moves						
Humor	HUM	.10	.17	.11	.07	.06
Anecdote/example	ANEX	.22	.18	.29	.24	.18
Graphic	GR	.52	.63	.63	.47	.37
Organization	ORG	.10	.12	.07	.09	.10
Emphasis	EMP	.05	.09	.03	.03	.04
Instructor–student interactions						
Instructor rhetorical questions	IRQ	.12	.10	.10	.12	.17
Instructor display questions[b]	IDQ	.36	.46	.39	.31	.29
Comprehension questions	ICQ	.07	.09	.07	.05	.06
Student novel question	SNQ	.04	.03	.02	.08	.02
Student comprehension question	SCQ	.11	.06	.20	.12	.09
Student response[c]	SR	.28	.36	.29	.25	.21
Student peer interactions	PI	.11	.10	.10	.11	.14
Cognitive engagement						
Problem solving	PS	.15	.11	.17	.17	.19
Creating	CR	.02	.03	.01	.04	0
Connecting to real world	CN	.25	.21	.30	.31	.20
Instructional technology						
Chalkboard	CB	.19	.01	.32	.15	.39
Overhead projector	OP	.08	.04	.19	.05	.07
PowerPoint	PP	.57	.86	.25	.56	.41
Clicker response system	CL	.10	.16	.08	.04	.14
Demonstrations	D	.02	.01	.05	.01	.04
Digital tablet/document camera	DT	.10	.11	.21	.03	.08
Movies	M	.03	.03	.03	.02	.02
Simulations	SI	.01	0	.02	0	.01

[a] 2^+ questions posed.
[b] Seeking new information.
[c] To instructor question.

data were analyzed to categorize instructors based on the length of time that they used a lecturing modality without any form of visible student engagement, as indicated by the TDOP codes. The results indicate that 61% of the sample ($n = 34$) lectured with no visible student engagement for periods of 20 minutes or less, 23% ($n = 13$) lectured for periods between 21

and 40 minutes, and 16% ($n = 9$) lectured for over 40 minutes. These data indicate that the majority of faculty in the study sample engaged in what is popularly known as a "straight lecture" for relatively brief segments of time within their class. That said, nine instructors were observed teaching in ways where students were passive for almost the entire class period.

Other important aspects of teaching include pedagogical strategies not tied to any particular teaching method including the use of anecdotes and examples (ANEX, 22%) and organizational markers (ORG, 10%). Similarly, faculty employ different approaches to interacting with students, particularly through the posing of questions, including rhetorical questions where students are not expected to answer (IRQ, 12%) or more open-ended questions, known as display questions that solicit specific information from students (IDQ, 36%). Different types of student cognitive engagement were also documented including making connections to the real world (CN, 25%) and a variety of instructional technologies were observed, especially PowerPoint slides (PP, 57%) and chalkboards (CB, 19%).

Variation by Disciplinary Group. This analysis also highlights variations among the disciplinary groups included in the study sample. While an extended analysis of these data is beyond the purview of this paper and has been conducted elsewhere (see Hora & Ferrare, 2013), it is worth pointing out certain differences across groups. For example, in regard to teaching methods, the biologists in the study sample engaged in more lecturing with premade visuals (LPV, 86%) and assessments (A, 17%) than other disciplines, whereas mechanical engineers and physicists used more lecturing with hand-made visuals (LHV, 52% and 49%, respectively), working through problems (WP, 26% and 19%), and desk work (DW, 12% and 12%) than the biologists and geoscientists. Other differences are apparent across other dimensions, including the relatively high incidences of effective pedagogical strategies such as organization (ORG, 12%) and emphasis (EMP, 9%) by biology instructors, the high rates of student comprehension questions (SCQ, 20%), and the use of digital tablets in mechanical engineering classes (DT, 21%). The use of tablets was particularly pronounced at one institution, which had previously invested in upgrading classroom technologies. Similarities across groups are also worth noting, such as the similar rates of certain teaching methods (SGW), instructor-student interactions (PI) and cognitive engagements (PS).

Variation by Course Structure. Given that teaching practices may be influenced by contextual factors such as course level and class size, Table 6 includes data grouped according to these variables. Besides highlighting another point of variability within the data, this analysis also underscores the important role that organizational contexts plays in shaping how instruction unfolds in the classroom as suggested by systems-of-practice theory.

In terms of differences observed according to course level, analysts in lower division classrooms observed less small group work (SGW, 9% and 15%, respectively), peer interactions (PI, 9% and 14%), and creating cognitive engagements (CR, 1% and 5%). Results such as these suggest that upper division courses tend to afford students more opportunities for active engagement with one another and the course material. Variation in teaching practices were also evident in classrooms of differing sizes, though in several cases no clear patterns were evident. For example, lecturing with premade visuals was observed at higher rates in courses with 26 to 100 students (LPV, 71%) and 200 or more students (82%) and at lower rates in courses with less than 25 students (48%) and between 100 and 199 students (55%). In other cases, however, the data suggest that the cognitive engagement of creating is

TABLE 6
Classroom Observation Data Using the TDOP by Discipline and Course Level

		Course Level		Class Size			
		Upper	Lower	<25	26–100	101–199	200+
Instructors		22	34	8	18	18	12
Total 2-minute intervals		1,030	1,477	344	828	821	521
Teaching methods							
Lecturing	L	.07	.06	.10	.06	.08	.02
Lecturing with premade visual	LPV	.60	.68	.48	.71	.55	.82
Lecturing with handmade visual	LHV	.33	.24	.36	.18	.42	.15
Lecturing with demonstration	LDEM	.04	.03	.04	.00	.07	.03
Socratic lecture[a]	SOC-L	.05	.02	.02	.14	.07	.02
Working through problems	WP	.06	.09	.03	.09	.08	.06
Small group work	SGW	.15	.09	.13	.12	.09	.13
Desk work	DW	.08	.06	.15	.02	.08	.05
Multimedia	MM	.04	.02	.02	.03	.03	.01
Assessment	A	.10	.13	.06	.05	.15	.19
Pedagogical moves							
Humor	HUM	.10	.11	.06	.06	.13	.18
Anecdote/example	ANEX	.21	.23	.27	.25	.21	.18
Graphic	GR	.49	.53	.35	.46	.57	.54
Organization	ORG	.08	.11	.09	.10	.08	.12
Emphasis	EMP	.03	.06	.03	.04	.04	.10
Instructor–student interactions							
Instructor rhetorical questions	IRQ	.07	.15	.07	.14	.11	.12
Instructor display questions[b]	ICQ	.35	.37	.34	.27	.40	.47
Comprehension questions	CQ	.06	.07	.03	.09	.06	.03
Student novel question	SNQ	.02	.05	.02	.04	.06	.03
Student comprehension question	SCQ	.13	.10	.11	.14	.12	.06
Student response[c]	SR	.28	.29	.31	.21	.30	.37
Student peer interactions	PI	.14	.09	.11	.10	.09	.15
Cognitive engagement							
Problem-solving	PS	.17	.14	.17	.14	.16	.17
Creating	CR	.04	.01	.00	.06	.00	.00
Connecting to real world	CN	.24	.28	.30	.30	.24	.21
Instructional technology							
Chalkboard	CB	.25	.14	.36	.17	.25	.00
Overhead projector	OP	.07	.09	.08	.05	.13	.04
PowerPoint	PP	.47	.63	.40	.60	.44	.81
Clicker response systems	CL	.08	.12	.00	.05	.14	.19
Demonstrations	D	.03	.02	.04	.00	.05	.01
Digital tablet/document camera	DT	.12	.09	.00	.06	.12	.19
Movies	M	.04	.02	.00	.04	.03	.02
Simulations	SI	.00	.01	.06	.00	.02	.00

[a]2+ questions posed.
[b]Seeking new information.
[c]To instructor questions.

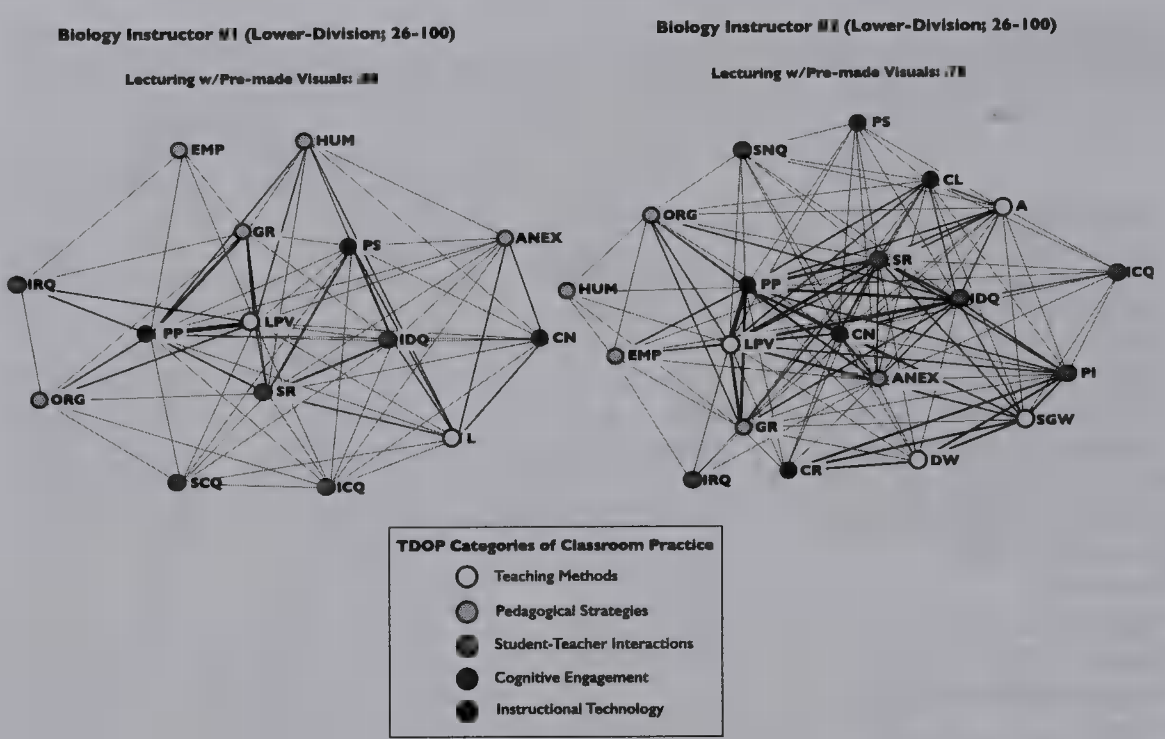


Figure 1. Co-occurence network graph for two biologists teaching similar courses.

most evident in classes with 26–100 students, and that the use of instructional technologies such as PowerPoint slides and clicker response systems becomes more prevalent as class size increases.

Social Network Analysis Graphs

While these data provide a fine-grained account of classroom practice across various settings, the act of decomposing teaching into singular variables such as LPV perpetuates the fiction that teaching can be adequately represented by such measures. Instead, a more accurate approach is to identify the relative intensity with which certain combinations of codes were observed throughout the class period. To illustrate how different aspects of classroom dynamics interact with one another over time, I used social network analysis to depict the TDOP data of two biology instructors at the same institution who were both teaching lower division classes that ranged in size from 26 to 100 students (see Figure 1).

The graphs depicted here illustrate two key points. First, these graphs clearly show that individual codes do not exist in a vacuum but instead often co-occur within the same 2-minute interval. The lines connecting the codes vary in thickness (on a scale of 1–5) depending on the number of times each pair of codes was observed in the same interval. As a result, the thicker lines indicate an increased frequency with which codes were observed during the same time interval. The codes are also positioned in ways that minimize the variation in line length, such that codes closer to the center of the graph tend to be those that are more frequently observed together. Ultimately, these images bring to life the advantages of the instructional systems-of-practice framework—how distinct dimensions of instructional practice interact with one another in different ways during the course of a 50-minute class.

Second, the graphs highlight the limitations with conceptions of teaching that collapse the dynamism and complexity of classroom teaching into singular descriptors or variables

such as "lecturing." In this case, both faculty exhibited a high degree of lecturing with premade visuals (LPV, 80% and 78%, respectively), yet their overall teaching practices as captured by the TDOP vary considerably. Instructor #1 relied on instructional episodes, which are interconnected modes of instruction that represent a distinct teaching "event" in the classroom, primarily involving lecturing with PowerPoint slides, with other more peripheral behaviors (e.g., using anecdotes) seen less frequently in her class. Instructor #2 also relied on lecturing with PowerPoint slides but demonstrated a far more diversified range of practices that included a variety of teaching methods, pedagogical strategies, and student–instructor interactions. Thus, the graphs demonstrate that while both faculty lectured for extensive periods of time, their teaching approaches and thus the learning environment for their students were quite different, even within similar organizational conditions.

What is the Prevalence and Nature of Active Learning Observed in These Classrooms?

Next, I address the primary question addressed in this paper—to what degree do results indicate the prevalence (or lack thereof) of different types of active learning as measured by categories in the DOLA framework? This question is answered using two different analytic techniques. First, code combinations indicative of active learning are presented according to disciplinary affiliation, course level, and class size. Second, to delve more deeply into how active learning modalities are being used in a specific context, social network graphs are presented that depict instructional practices in upper and lower division courses.

Code Frequencies

The data indicate that the three different categories for active learning suggested by the DOLA framework were observed in varying degrees across the study sample. The "being active" category was most frequently observed, particularly when students responded to instructor questions (SR, 28% of all 2-minute intervals) followed by students engaging in the problem solving modes of cognition (PS, 15%). Much less frequently observed was the "being constructive" category of active learning, which included students asking novel questions (SNQ, 4%) and engaging in creative modes of cognition (CR, 2%), as well as the "being interactive" category of active learning that included both peer instruction and the creating cognitive engagement mode (PI–CR, 2%). In addition, analysts observed variations across disciplinary affiliation and course context (i.e., course level and class size) as seen in Table 7.

Variation by Disciplinary Group. In looking at the results across disciplines, the "being active" category as measured by the TDOP codes of SR and PS was observed the most frequently. The code for student responses to questions was observed the most in biology classrooms (SR, 36% of all 2-minute intervals), followed by physics (29%), geoscience (25%), and then mechanical engineering (21%). The frequencies for problem solving code (PS) was observed in the reverse order. Next, the two categories that have been found to be more effective than the being active modality – that of "being constructive" and "being interactive" – were observed less frequently. The being constructive category involves students either asking novel questions (SNQ) or being actively engaged in open-ended tasks (CR). For both of these codes, the highest rates were in geoscience courses (8% and 4%, respectively). Overall, however, this category of active learning was observed infrequently across the sample relative to the being active category. This was also the

TABLE 7
Description of TDOP Code Combinations Used to Capture Elements of the Differentiated Overt Learning Activities (DOLA) Framework (Chi & Wylie, 2014)

		Disciplinary Group				Course Level		Class Size			
DOLA Category	All Sample	Biology	Physics	Geoscience	Mechanical Engineering	Lower Division	Upper Division	<25	26–100	101–199	200+
Being active											
Student response (SR)	0.28	0.36	0.29	0.25	0.21	0.29	0.28	0.31	0.21	0.30	0.37
Problem-solving (PS)	0.15	0.11	0.17	0.17	0.19	0.14	0.17	0.17	0.14	0.16	0.17
Being constructive											
Student novel question (SNQ)	0.04	0.03	0.02	0.08	0.02	0.05	0.02	0.02	0.04	0.06	0.03
Creating (CR)	0.02	0.03	0.00	0.04	0.01	0.01	0.04	0.00	0.06	0.01	0.00
Being interactive											
Creating and peer interactions (CR and PI)	0.02	0.02	0.00	0.04	0.01	0.01	0.04	0.00	0.05	0.01	0.00

case with the category theorized to be the most effective mode of active learning—that of "being interactive"—which requires both students' active construction of knowledge *and* interaction with peers.

Variation by Course Context. Results for different course levels and class sizes also demonstrate variations in modes of active learning. For the being active mode, the highest rates of students responding to questions were observed in classes with 200 or more students (SR, 37%) and in classes with less than 25 students (31%). This result indicates that either very small or relatively large classes (more than 100) facilitate higher rates of students actively responding to questions, whereas in the mid-size classes (26–100) lower rates were observed (21%). The problem solving mode of cognitive engagement was observed with less variation across course level and class size (from 14 to 17%). Students asking novel questions, one of the indicators for the "being constructive" mode, was observed more frequently in lower division (SNQ, 5%) than in upper division courses (2%), and most often in classes with 101–199 students (6%). These data indicate that students are asking their instructors original questions with greater regularity in lower division courses and in relatively large classes. The other code for this category, that of the creating cognitive engagement, was observed more often in upper division (CR, 4%) than in lower division courses (1%), and most frequently in classes with 26–100 students (6%). Finally, in regard to the "being interactive" category of active learning, the course contexts in which this type of student activity was observed were primarily those of upper division courses (PI-CR, 4%) and in classes with 26–100 students (5%). These results indicate that students are engaging in open-ended problem solving tasks with their peers often in upper division courses and in classes of a particular size (26–100 students).

Social Network Analysis Graphs

Next, I used social network analysis techniques to depict the TDOP data across course levels for illustrative purposes. In the figures presented in this section, a single code representing one of the DOLA categories was the starting point of the analysis (e.g., SR and CR), and all other codes that were observed with these indicators of active learning for at least 10 2-minute intervals were included in the graph (i.e., an ego network). Thus, every code included in these graphs co-occurred with one of the indicators of active learning. A large arrow indicates the primary code of interest in each graph.

Active Learning in Lower Division Courses. To explore the prevalence of active learning in lower division courses, I created two co-occurence network graphs. The first graph highlights the SR code that indicates the "being active" modality (see Figure 2).

This graph is rather complex given that 25 different dimensions of teaching were observed co-occurring with students responding to questions. The codes that most frequently co-occurred with SR in a given 2-minute interval included instructors posing display questions (IDQ, 411 of a total of 1484 intervals), lecturing with premade visuals (LPV, 313 intervals), PowerPoint slides (PP, 287 intervals), and the use of graphics (GR, 166 intervals). The results highlight a commonly observed instructional episode where in the midst of a period of verbal exposition with PowerPoint slides, an instructor would interrupt his or her lecture with one or more questions posed to the class.[7] This suggests that in these courses the

[7]It is important to note that these instructional episodes did not occur identically across all observed class periods, as their exact duration and enactment necessarily varies by instructor and situation.

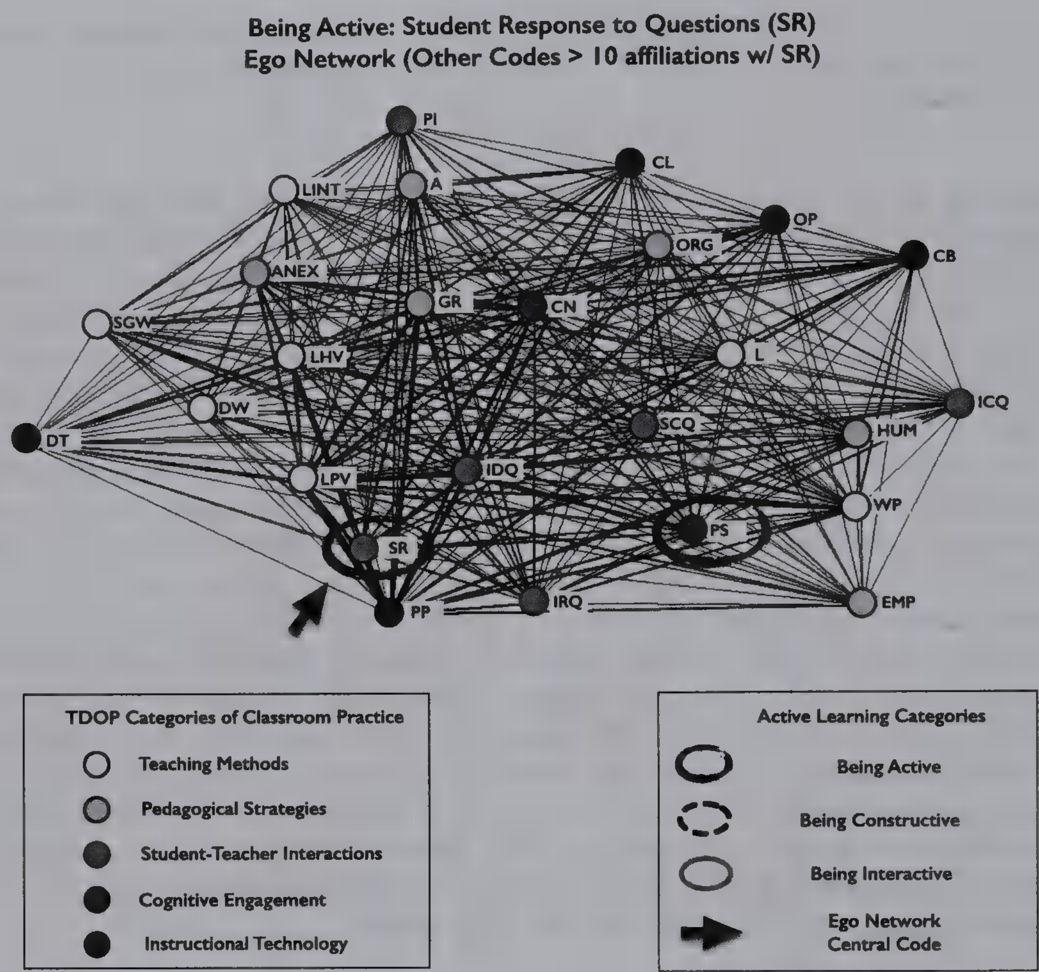


Figure 2. Co-occurence network graph for the "being active" mode in lower division courses as indicated by the "student-response" (SR) code.

"being active" modality as indicated by the SR code is tightly linked to, if not dependent upon, lecturing with specific types of instructional technologies (i.e., PowerPoint slides).

Another graph was created for the other indicator of "being active"—that of PS—but in the interests of space, and since this graph differed from the SR graph only by the exclusion of six codes (L, LPV, SOC-L, ORG, EMP, and IRQ), this graph is not depicted here. The analysis revealed, however, that 19 different dimensions were observed co-occurring with students engaging in PS, with the most frequently observed being instructors' display questions (IDQ, 137 intervals), PowerPoint slides (PP, 108 intervals), peer interactions (PI, 98 intervals), student response to questions (SR, 97 intervals), and small group work (SGW, 95 intervals). These results suggest that the instructional episodes associated with problem solving include questions posed during lectures that feature PowerPoint slides and student responses to these questions either alone or in small groups. Other codes associated with PS that were observed at lower rates of co-occurrence, including clickers (CL, 76 intervals) and desk work (DW, 72 intervals) shed more light on the instructional episodes that involve student problem solving.

The final graph depicts instances where one of the codes signifying the "being constructive" active learning modality—that of students posing novel questions to the instructor (SNQ)—was observed in lower division classrooms (see Figure 3).

The other code for the being constructive mode is not depicted because it was not observed (i.e., creating, CR). Consequently, the "being interactive" modality is not depicted in a graph

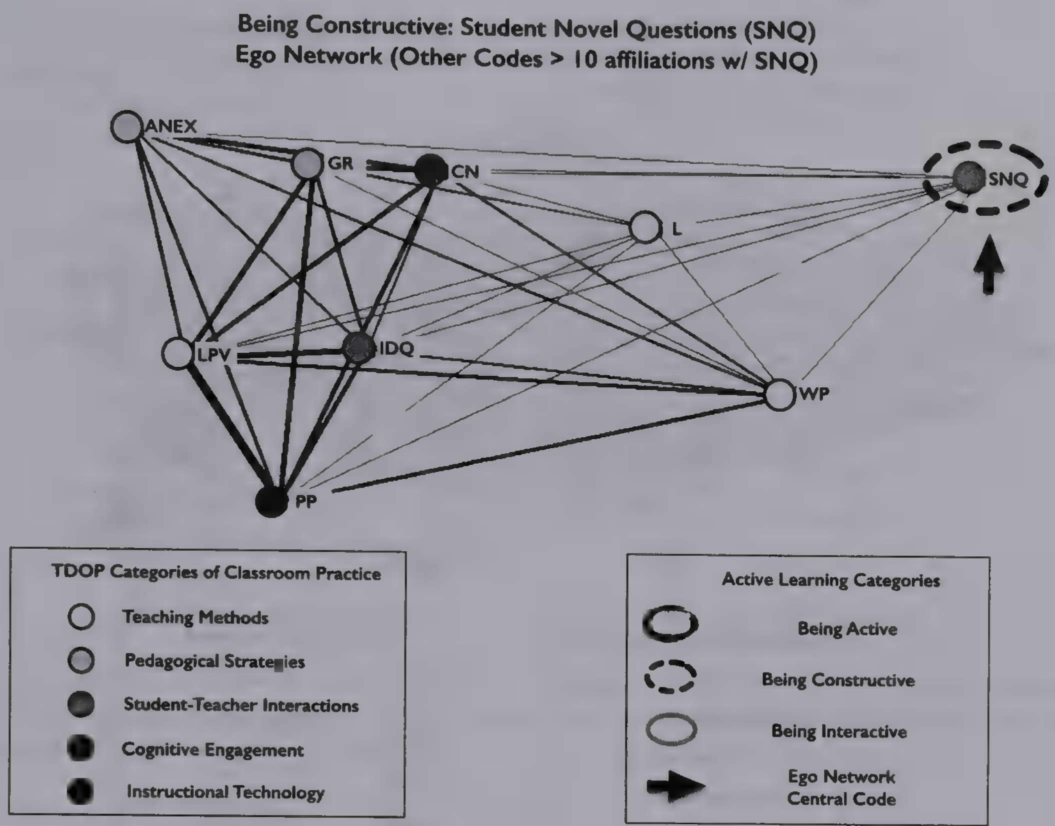


Figure 3. Co-occurence network graph for the "being constructive" mode in lower division courses as indicated by the "student's asking novel questions" (SNQ) code.

for lower division courses. For students posing novel questions, however, eight other TDOP codes were observed with the SNQ code, including lecturing with premade visuals (LPV, 41 intervals), PowerPoint slides (PP, 40 intervals), lecturing (L, 31 intervals), and instructor's working out problems (WP, 14 intervals). It is important to note that in contrast to the preceding graphs, the central code in the ego network is not at the center of the graph but instead occupies a space on the periphery. This indicates that episodes involving SNQ are not frequently observed in lower division classes.

Active Learning in Upper Division Courses. Next, I report findings from the analysis of active learning modalities observed in upper division courses. The first graph highlights the SR code that indicates the "being active" modality (see Figure 4).

This graph depicts the 26 codes that co-occurred with SR. The centrality of the SR code is immediately apparent in the graph, with several thick, dark lines connecting it to other dimensions of instruction. Additionally, a cluster of other active learning modalities is evident at the right of the graph (i.e., being interactive and being constructive), suggesting that a more peripheral set of practices exists at this course level in addition to the core modality of being active. The codes that most frequently co-occurred with SR in a given 2-minute interval included instructors posing display questions (IDQ, 263 of a total of 1030 intervals), lecturing with premade visuals (LPV, 184 intervals), PowerPoint slides (PP, 148 intervals), and the use of graphics (GR, 109 intervals). The results are identical to incidences of SR in lower division courses, albeit with different rates of co-occurrence. This similarity suggests that instructors across course levels commonly interrupt periods of verbal exposition with PowerPoint slides by posing one or more questions to the class.

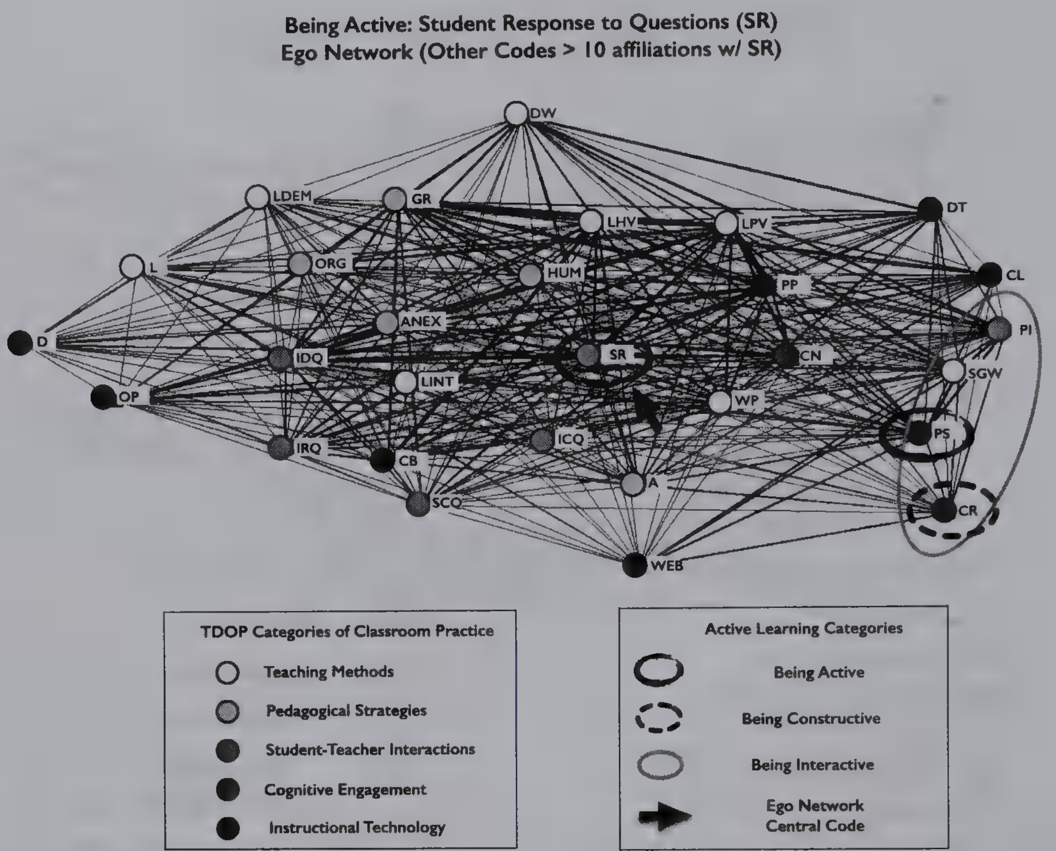


Figure 4. Co-occurence network graph for the "being active" mode in upper division courses as indicated by the "student-response" (SR) code.

In the interests of space, the graph for the upper division PS code, also indicative of the "being active" modality, is not included in this paper. This analysis did indicate that the PS graph only differed from the SR graph by the exclusion of eight codes (LDEM, SOC-L, IRQ, ICQ, SCQ, CN, OP, and D) and the presence of two codes (CR and WEB) not in the previous graph. The analysis revealed 20 different dimensions observed co-occurring with students engaging in PS, with the most frequently observed being small group work (SGW, 127 intervals), student peer interactions (PI, 110 intervals), instructor's display questions (IDQ, 105 intervals), and lecturing with premade visuals (LPV, 85 intervals). These results suggest that the instructional episodes associated with PS at the upper division level include students working in small groups with their peers on questions posed by the instructor. The instructional episode that links PS to SGW and IDQ as described for lower division courses is also evident here because SGW and IDQ also regularly co-occur (86 intervals). Furthermore, it is notable that many of the questions that spark PS are delivered from a lecturing context—that is, a period of lecturing often preceded instances of student problem solving.

The next figure includes two smaller graphs that depict both the "being constructive" and the "being interactive" modalities (see Figure 5).

In the first graph on the left, one of the indicators for the "being constructive" modality—that of student's posing novel questions (SNQ)—is depicted as the primary code in the graph. In this graph, only three other codes were observed co-occurring with students posing novel questions, with the most frequently observed being lecturing with premade visuals (LPV, 19 intervals), PowerPoint slides (PP, 13 intervals), and the use of graphics (GR, 10 intervals).

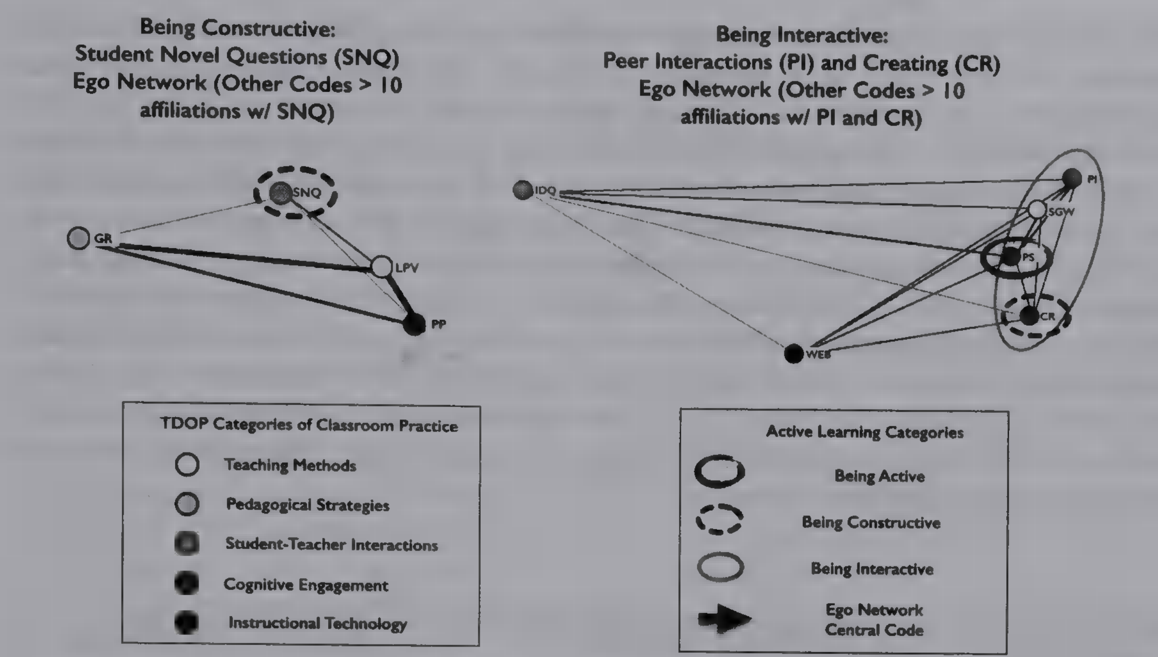


Figure 5. Co-occurence network graph for the "being constructive" mode in upper division courses as indicated by the "student's asking novel questions" (SNQ) code and the "being interactive" mode as indicated by the "peer interaction" (PI) and "creating" (CR) codes.

These results are similar to those observed in lower division courses, where students were observed interrupting instructor's Powerpoint lectures with novel questions.

In the second graph on the right side of Figure 5, the two indicators comprising the "being interactive" modality are shown. No graph is shown for the creating (CR) mode on its own, which is an indicator for "being constructive," because the networks for both CR alone and CR and PI together are identical. This co-occurence network graph depicts all TDOP codes that co-occurred with both CR and PI in the same 2-minute interval. These codes included small group work (SGW, 127 intervals with PI and 40 intervals with CR), instructors posing display questions (IDQ, 67 intervals with PI and 12 with CR), problem solving (PS, 110 intervals with PI and 31 with CR), and Web site use in the classroom (WEB, 33 intervals with PI and 32 with CR). These results show instructional episodes that were observed in classes where instructors posed display questions often focused on open-ended tasks or problems that initiated small group-work involving peer-to-peer interactions. In one case, students were directed to use a Web site as part of a problem solving activity.

DISCUSSION

The overall goal of this study was to advance an approach for studying teaching in naturalistic settings, with the specific aim of demonstrating how active learning modalities can be documented and described in postsecondary classrooms. In this paper, the "problem" of identifying whether active learning techniques are being used, which is of great interest to policymakers and researchers alike (e.g., PCAST, 2012; AAAS, 2012), was reframed from one that involved identifying the prevalence (or absence) of coarse descriptions of teaching to a more nuanced and comprehensive view of instructional practice. The latter formulation of the problem represents an entirely different proposition than the former. Instead of assuming that the fundamental question of understanding the relationship between teaching and learning in *all* settings (both controlled and naturalistic) was a settled empirical question, such that the mere identification of "lecturing" would capture the entirety of an instructor's

practice and subsequent student engagement (Freeman et al., 2014; Wieman, 2014), the problem is rather one of first attempting to describe the nature of teaching itself as a scientific problem on its own merit. Indeed, by demonstrating that modes of active learning are often embedded within PowerPoint lectures and that small group work exercises are not synonymous with constructivist activities, the evidence presented in this paper as well as by other researchers (e.g., Chi & Wylie, 2014) suggests that the research program of understanding teaching as a complex behavioral phenomena in its own right is in the early stages. As such, the field of science education would benefit from research instruments that can detect, with precision, what actually happens in real-world classrooms, something that existing surveys and classroom observation protocols are not designed to do. In the remainder of this paper, I elaborate on key findings from the observation of 56 science and engineering instructors using the TDOP, and outline implications for research, practice, and policy that follow from the results.

Prevalence of Instructional Practices in General: The Central Role of Lecturing

First, the results shed light on the types of instructional practices that 56 science and engineering instructors are using in their courses. In terms of single TDOP codes, the most frequently observed included those pertaining to different forms of verbal exposition (i.e., lecturing), particularly in conjunction with premade visuals (LPV, observed in 64% of all 2-minute intervals) and handmade visuals (LHV, 27%). While caution must be taken when isolating these data from their role as part of an interactive system of tools and practices, they do shed light on the prevalence of individual aspects of classroom dynamics. For example, the data speak to the widespread use of particular types of representations and media (e.g., premade visuals such as PowerPoint slides) that faculty use as part of their lecturing.

Specifying distinct types of lecturing according to the different representational modes used in the classroom is also important because such artifacts mediate learners' experiences with course material and the disciplinary processes and communities in which they are embedded (Brown et al., 1989; Lattuca, 2002). The tools and technologies used while lecturing also effectively frame and organize knowledge in particular ways that offer viewers a circumscribed range of potential responses or actions (Greeno, 1998), and which may ultimately reorganize observers' mental functioning (Hawkins & Pea, 1987). While the data reported in this paper are not designed to advance claims on the types of cognitive activities afforded by viewing a series of PowerPoint slides relative to writing on a chalkboard, research on relationships among the use of varied representations, technology, and student learning should be incorporated into debates about instructional reform (e.g., Mayer, 2011).

Furthermore, because 34 instructors (61%) lectured for relatively brief periods (less than 20 minutes), the data indicate that a majority of the faculty in the study were not engaged in extensive lecturing that lacked any interactions with students. Indeed, some instructors also interspersed questions, small group work, and other activities throughout their "lectures." This particular approach to lecturing was evident in the case of the biology instructor featured in Figure 1. Ultimately, these data demonstrate that in practice, lecturing can be a rather complex teaching practice that does not conform to the oft-cited caricature whereby the instructor speaks, with no effort or opportunity for student engagement or activity, for an entire class period. While nine instructors in the study certainly did enact the common notion of the "straight lecturer" as a teacher-centered performance with little care or attention toward engaging students, most did not. Indeed, in some cases, it appeared

that lecturing was used as a pedagogical device to support other, more engaged modes of instruction.

It also cannot be immediately assumed that the dominance of lecturing automatically leads to inferior instruction and poor student learning (see Schwartz & Bransford, 1998; Crouch & Mazur, 2001; Miller et al., 2008). Thus, the pertinent question becomes one of the amount and timing of lecturing and whether an underlying pedagogical rationale guides its use. One of the key questions facing the field, then, is how to use these "times for telling" in a way that is pedagogically sound, well executed, and serves to set up robust active learning modalities (Schwartz & Bransford, 1998). That said, while the specific duration and type of lecturing that is the most pedagogically beneficial remains an open question, the weight of the evidence regarding the importance of active learning does suggest that there is room for greater incorporation of nonlecture activities in some classrooms—particularly in the case of those nine faculty who lectured for over 41 minutes with nary a question or activity.

Finally, in regard to how the data compare with the extant literature on the prevalence of lecturing, such direct comparisons are not useful given the fact that it is impossible to determine what is meant by lecturing in many surveys and whether respondents understand the term in the same manner (Porter, 2011). However, some tentative conclusions may be drawn. While the HERI Faculty Survey (Hurtado et al., 2012) does not define what is meant by the term "extensive lecturing," the 63% of respondents who indicated this option as a regularly used teaching method is similar to the results for the use of premade visuals using the TDOP instrument (observed in 64% of all 2-minute intervals), but varies from results reported here regarding the prevalence of long periods of lecturing (e.g., 40^{+} minutes). This result suggests the possibility that survey respondents may interpret the phrase "extensive lecturing" not in terms of consecutive minutes of use but regarding its overall prevalence within a class. In any case, the problems inherent in such comparisons highlight the fact that despite calling for the field to operationally define "lecturing" over 20 years ago, Schonwetter's (1993) call has not yet been answered.

Prevalence of Active Learning Modalities: Variation Within Instructional Activities

Second, the results also shed light on the prevalence of active learning in the postsecondary classroom. Using indicators of active, constructive, and interactive types of instructional activities as posited by the DOLA framework (Chi & Wylie, 2014), the evidence shows that the most commonly observed type of active learning involved the "being active" category. This most often involved students responding to instructor's questions (SR, observed in 28% of all 2-minute intervals) followed by students engaging in problem solving modes of cognition (PS, 15%). These types of active learning frequently involved instructional episodes that included instructors posing questions that sought new information, either verbally or through electronic means (e.g., clicker response systems), often during a period of lecturing with PowerPoint slides. It is important to note that this category of active learning is at the lowest end of the DOLA framework in terms of efficacy in supporting student learning, and while it is more effective than a lesson where students are completely passive, it is less effective than instruction that facilitates students to be "constructive" or "interactive" (Chi & Wylie, 2014). Ultimately, the results suggest that the "being active" mode is a common instructional approach in postsecondary classrooms, achieved primarily through questions posed to students in the midst of a PowerPoint lecture, whereas other types of active learning are far less common.

In comparing these data with the literature, keeping the caveats regarding response process issues and comparability in mind, the results are both consistent and contradictory.

First, the HERI survey showed that 47% of faculty used small groups and 61.5% used class discussions (Hurtado et al., 2012), whereas the results reported here indicate smaller use of these methods (11% and 0%, respectively). In the Henderson and Dancy (2009) survey, 13.9% of the respondents reported using Peer Instruction and 13.7% reported using group work. While no TDOP codes reflect Peer Instruction in strict terms, the use of clickers (CL, 8%) and small group work (SGW, 11%) capture key facets of the strategy and similar to findings in other studies. What is most evident from the comparison of these survey results to the findings reported in this paper is the fundamentally different perspective on teaching offered by the instructional systems-of-practice framework and most survey instruments. Instead of distilling instruction in general and active learning in particular down to single descriptors, which are often assumed to be synonymous with high-quality student cognitive engagement, active learning modalities are seen as being composed of a configuration of distinct yet interrelated dimensions of teaching. The limited capacity of surveys to capture the prevalence (and quality) of active learning is perhaps most evident in the example of small group work, a widely encouraged classroom activity (e.g., Handelsman et al., 2004; PCAST, 2012). However, it cannot be assumed that asking students to work in groups automatically translates into an interactive and effective learning experience (Chi & Wylie, 2014).

The data reported in this paper support this idea by revealing that small group work activities vary in terms of associated cognitive demands between lower division and upper division classes. In lower division classes, small group work was more often associated with closed-ended problem solving modes of cognition (PS, observed with SGW in 95 of 1477 total 2-minute intervals) than open-ended creative modes of cognition (CR, observed with SGW in 0 of 1477). Similarly, for upper division classes, small group work was more often associated with problem solving (PS, observed with SGW in 127 of 1030) than creative modes of cognition (CR, observed with SGW in 40 of 1030). These findings are consistent with prior research that small group activities can be implemented in very different ways (Turpen & Finkelstein, 2009), and highlight the fact that without differentiating among types of student cognitive engagement (e.g., CR and PS), relying solely on whether or not small group work is taking place results in an incomplete and potentially misleading account of active learning in the classroom (e.g., Smith et al., 2013).

Insights Into Variations by Disciplinary Group and Course Characteristics

Finally, the results indicate variations in both teaching practices in general and active learning modalities in particular across different disciplinary groups and course contexts (i.e., course level and class size). While an extended analysis of the nature of and implications for these differences are beyond the purview of this paper, it is worth noting that such variations are consistent with a theory of practice that emphasizes the dynamic interactions between context and activity (Halverson, 2003; Spillane et al., 2001). In particular, the interactions between organizational structure and teaching practices are evident in the different rates of the "being constructive" modality, where students were most frequently observed asking novel questions in lower division (SNQ, observed in 5% of all 2 minute intervals) courses, and in classes with 101–199 students (SNQ, 6%). These and other points of variation suggest that the level of the course as well as structural affordances in the classroom itself may play an important role in shaping both the instructional decisions made by instructors and students' subsequent cognitive activity.

Limitations

There are several limitations to the TDOP and the data reported in this study. First, being a descriptive instrument, the TDOP does not shed any light on whether or not any particular teaching behaviors are being used to good effect. Additional sources of data regarding the efficacy of certain instructional practices (e.g., assessments of student learning) are needed to arrive at any estimation of instructional quality. Second, one of the major limitations with observation-based data is that it relies on the observer to infer whether or not a particular behavior has occurred. While this limitation is less problematic with regard to capturing discrete, easily interpreted phenomenon (e.g., the use of instructional technology), it becomes a significant issue if the intent of an observation is to estimate phenomena such as potential student cognitive engagement. Thus, there remains a certain degree of error associated with any given code frequency. Third, while the training described in this paper was rather extensive, only 2% of the dataset was used to establish IRR. In the future, training should include more videotaped lectures to test IRR (e.g., 10–15), though trade-offs with the increased time for training should be considered. Finally, limitations to the study reported in this paper include the self-selected nature of the sample, the lack of observations conducted throughout the course of a term, and the absence of data on laboratory or discussion sections.

Implications for Research, Policy, and Practice

Based on the evidence reported in this paper, I argue that the field of science education stands to benefit from a more careful discussion of teaching in terms of research, policy, and practice.

Research

In attempting to study, understand, and measure classroom teaching as a behavioral phenomenon itself it is clear that singular, decontextualized metrics and binary categorization schemes are insufficient. If the field were to think of the problem in terms of biological classification, it is as if researchers were sometimes focusing solely at the level of classes, while ignoring taxa such as genera or species that would reveal more fine-grained and subtle distinctions among different types of instruction. The instructional systems-of-practice approach and the TDOP represent a step in this direction, but future research should continue to identify, in more precise terms than at present, the types of teaching practices, active learning modalities, and student behaviors that occur in real-world classrooms. Towards this end, my research group is actively field-testing new sets of TDOP codes to capture more nuanced features of active learning techniques as well as instruction that conveys what are known as "21st century competencies" important for life and work (e.g., self-regulation) (Pellegrino & Hilton, 2012). Indeed, some researchers are actively using the basic architecture of the TDOP to develop new methods for studying aspects of active learning in the classroom (Lund et al., 2015).

In addition, researchers should focus on better understanding student behavior and subsequent cognitive demands. As Good and Brophy (2000) noted in their discussion of effective teaching in K-12 schools:

> Observers often try to reduce the complexity of classroom coding by focusing their attention exclusively on the teacher . . . but it is misplaced emphasis. The key to thorough classroom observation is student response. If students are actively engaged in worthwhile

> learning activities, it makes little difference whether the teacher is lecturing, using discovery techniques, or using small-group activities for independent study. (p. 47)

Additionally, in studying whether and how instructors are adopting active learning techniques, researchers should expand their inquiries beyond the classroom. Factors such as careful attention to course design as a whole and how instruction, assessment, and student assignments can work together in a complementary manner to provide students with a rich learning experience are just as important to the provision of high-quality learning opportunities (Freeman, Haak, &Wenderoth, 2011). Furthermore, given increasing evidence regarding which study strategies are most effective in facilitating student learning (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), researchers should pay as much attention to student study habits "in the wild" as is currently being focused on what faculty do in the classroom.

Practice

One of the most promising uses for descriptive data is that of supporting professional development efforts. A critical feature of professional development is the provision of credible and detailed feedback for instructors that can spark critical reflection (Chism, 2007). Yet data about teaching, beyond ubiquitous end-of-term student evaluations, are often in short supply in many colleges and universities, and faculty are often left with little data upon which to evaluate their teaching effectiveness (Gormally, Evans, & Brickman, 2014). Such reflection is a cornerstone to the ongoing development of professional expertise (Schön, 1983), and I suggest that the TDOP can be useful in providing data for these purposes, especially because faculty may be more responsive to and less threatened by the results in contrast to evaluative protocols such as the RTOP (Yon, Burnap, & Kohut, 2002). This thesis is currently being tested in a field-based study on faculty reactions to data from the TDOP and student evaluations as a new type of formative feedback (see http://tpdm.wceruw.org).

Policy

Unfortunately, the growing evidence supporting the effectiveness of active learning approaches and the binary categorization scheme used to characterize teaching has coalesced into pronouncements that lecturing is the "pedagogical equivalent of bloodletting" and is synonymous with an "inferior education" (Weiman, 2014, p. 8320). As such, the current rhetoric perpetuates the mistaken notion that the term "lecturing" refers to a distinct type of instructional practice that is unequivocally indicative of a specific type of student cognitive engagement, instead of one that may vary in a multitude of ways and actually be used in conjunction with active learning techniques. Such an approach can easily alienate instructors "in the field," as practitioners grow to resent the perceived imposition of particular philosophies or opinions about what constitutes effective or high-quality teaching from external authorities (Henderson & Dancy, 2008; Yon et al., 2002). Indeed, extensive research on reform implementation indicates that recipients of an intervention are most receptive when the messaging accompanying the innovation, not to mention the innovation itself, is closely aligned with the existing practices, cultural traditions, and beliefs of the population, instead of being in direct opposition to them (Rogers, 1995; Spillane et al., 2002).

Ultimately, I speculate that reforms that promote active learning may be more readily adopted, or at least seriously considered, if the messaging used by advocates reflects an understanding of the actual teaching practices used by faculty in their daily work—which

the lecturing versus interactive teaching framework fails to do. Indeed, given that a majority of faculty in the study reported in this paper relied on some form of verbal exposition in their classes, it is possible that suggesting slight modifications to the lecturing method may be a more promising approach than calling for the outright transformation of an instructor's entire pedagogical approach (Martin & Ramsden, 1993). Thus, it may not be a matter of eliminating lecture from one's pedagogical toolkit, but instead altering one's lecturing approach to have more of a deliberate pedagogical purpose that engages students in their own learning. An extensive amount of research is underway on this point (e.g., Walker, Cotner, Baepler, & Decker, 2008), and I suggest that the field will be better served in making minor yet influential adaptations to how verbal exposition is used in the classroom, rather than advocating for its complete and utter elimination.

I would like to thank the anonymous reviewers and Jana Bouwma-Gearhart for their feedback on earlier versions of this paper. The research reported here was supported by the National Science Foundation under award DUE#1224624. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Adams, C. (2006). PowerPoint, habits of mind, and classroom culture. Journal of Curriculum Studies, 38(4), 389–411.

American Association for the Advancement of Science. (2012). Describing and measuring undergraduate STEM teaching practices: A report from a national meeting on the measurement of undergraduate STEM teaching, December 17–19, 2012. Washington, DC: Author.

Blumenfeld, P., Kempler, T., & Krajcik, J. (2006). Motivation and cognitive engagement in learning environments (pp. 475–488). New York, NY: Cambridge University Press.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). UCINET for Windows: Software for social network analysis. Harvard, MA: Analytic Technologies.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). How people learn: Brain, mind, and school. Washington, DC: National Research Council.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32–42.

Cash, A. H., Hamre, B. K., Pianta, R. C., & Meyers, S. S. (2012). Rater calibration when observational assessment occurs at large scales: Degree of calibration and characteristics of raters associated with calibration. Early Childhood Research Quarterly, 27(3), 529–542.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science, 1(1), 73–105.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219–243.

Chism, N. V. N. (2007). Peer review of teaching: A sourcebook (2nd ed.). Bolton, MA: Anker.

Clark, R. M., Norman, B. A., & Besterfield-Sacre, M. (2014). Preliminary experiences with "flipping" a facility layout/material handling course. In Y. Guan & H. Liao (Eds.), Proceedings of the 2014 Industrial and Systems Engineering Research Conference, May 28–June 3. Montreal, Canada.

Coburn, C. E. (2001). Collective sense making about reading: How teachers mediate reading policy in their professional communities. Educational Evaluation and Policy Analysis, 23(2), 145–170.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. Educational Researcher, 32(6), 3–12.

Code, W., Piccolo, C., Kohler, D., & MacLean, M. (2014). Teaching methods comparison in a large calculus class. ZDM Mathematics Education 46(4), 1–13.

Cohen, D. K., & Ball, D. L. (1999). Instruction, capacity, and improvement. Consortium for Policy Research in Education Rep. No.RR-43. Philadelphia: University of Pennsylvania, Graduate School of Education.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. American Journal of Physics, 69, 970–977.

Danielson, C. (2013). The framework for teaching evaluation instrument (2013 edition). The Danielson Group.

Derting, T., Williams, K. S., Momsen, J. L., & Henkel, T. P. (2011). Education research: Set a high bar. Science, 333, 1220.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. Science, 332(6031), 862–864.

Duch, B. J., Groh, S. E., & Allen, D. E. (Eds.). (2001). The power of problem-based learning: A practical" how to" for teaching undergraduate courses in any discipline. Sterling, VA: Stylus Publishing.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public Interest, 14(1), 4–58.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. BioScience, 61(7), 550–558.

Finelli, C. J., Daly, S. R., & Richardson, K. M. (2014). Bridging the research to practice gap: Designing an institutional change plan using local evidence. Journal of Engineering Education, 103(2), 331–361.

Franklin, S. V., & Chapman, T. (2012). Diversity of faculty practice in workshop classrooms (Vol. 1513, pp. 130–133). Philadelphia, PA: Physics Education Research Conference.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415.

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. CBE-Life Sciences Education, 10(2), 175–186.

Good, T., & Brophy, J. (2000). Looking in classrooms. (8th ed.). New York: Longman.

Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. CBE-Life Sciences Education, 13(2), 187–199.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. American Psychologist, 53(1), 5–26.

Guarino, C., & Tracy, B. (2012). Review of gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Boulder, CO: National Educational Policy Center.

Halverson, R. (2003). Systems of practice: How leaders use artifacts to create professional community in schools. Educational Policy Analysis Archives, 11(37), 1–35.

Halverson, R. R., & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. Educational Administration Quarterly, 42(4), 578–619.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Scientific teaching. Science, 304(5670), 521–522.

Handelsman, J., Miller, S., & Pfund, C. (2007). Scientific teaching. New York: W.H. Freeman.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. Journal of Research in Science Teaching, 24(4), 291–307.

Henderson, C., & Dancy, M. H. (2008). Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. American Journal of Physics, 76(1), 79–91.

Henderson, C. R., & Dancy, M. H. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. Physical Review Special Topics–Physics Education Research, 5, 020107.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107.

Hora, M. T. (2014a). Limitations in experimental design mean that the jury is still out on lecturing. Proceedings of the National Academy of Sciences, 111(30), 3024.

Hora, M. T. (2014b). Exploring faculty beliefs about student learning and their role in instructional decision-making. The Review of Higher Education, 38(1), 37–70.

Hora, M. T., & Ferrare, J. (2013). Instructional systems of practice: A multi-dimensional analysis of math and science undergraduate course planning and classroom teaching. The Journal of the Learning Sciences, 22(2), 212–257.

Hora, M. T., & Ferrare, J. (2014). Re-measuring postsecondary teaching: How singular categories of instruction obfuscate the multiple dimensions of classroom practice. Journal of College Science Teaching, 43(3), 36–41.

Hurtado, S., Eagan, K., Pryor, J. H., Whang, H., & Tran, S. (2012). Undergraduate teaching faculty: The 2010–2011 HERI faculty survey. Los Angeles, CA: Higher Education Research Institute, UCLA.

Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate scores. MET project policy and practice brief. Bill & Melinda Gates Foundation.

Kane, M. T. (2001). Current concerns in validity theory. Journal of Educational Measurement, 38(4), 319–342.

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. Review of Educational Research, 72(2), 177–228.

Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. Learning and Instruction, 7, 255–275.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86.

Lattuca, L. R. (2002). Learning interdisciplinarity: Sociocultural perspectives on academic work. The Journal of Higher Education, 73(6), 711–739.

Lave, J. (1988). Cognition in practice: Mind, mathematics and culture in everyday life. Cambridge, UK: Cambridge University Press.

Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. CBE-Life Sciences Education, 14(2), 1–12.

MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. The Physics Instructor, 40, 479.

Martin, E., & Ramsden, P. (1993). An expanding awareness: How lecturers change their understanding of teaching. Research and Development in Higher Education, 15, 148–155.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? Educational Evaluation and Policy Analysis, 21(1), 29–45.

Mayer, R. E. (2011). Instruction based on visualizations. In R. E. Mayer & P. A. Alexander (Eds.), Handbook of research on learning and instruction (pp. 427–445). New York: Routledge.

Mazur, E. (2009). Farewell, lecture. Science, 323(5910), 50–51.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt. Learning activities for effective instruction in engineering classrooms. Journal of Engineering Education, 102(3), 346–374.

Miller, S., Pfund, C., Pribbenow, C. M., & Handelsman, J. (2008). Scientific teaching in practice. Science, 322(5906), 1329–1330.

Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. Journal of Educational Psychology, 75, 138–149.

National Research Council. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. Washington, DC: National Academies Press

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. Research in the Teaching of English, 25(3), 261–290.

Osthoff, E., Clune, W., Ferrare, J., Kretchmar, K., & White, P. (2009). Implementing Immersion: Design, professional development, classroom enactment and learning effects of an extended science inquiry unit in an urban district. Madison, WI: University of Wisconsin-Madison.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), Distributed cognitions (pp. 47–87). New York: Cambridge University Press.

Perry, R. P., & Smart, J. C. (Eds.). (1997). Effective teaching in higher education: Research and practice. New York: Agathon Press.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. Educational Researcher, 38(2), 109–119.

Porter, S. R. (2011). Do college student surveys have any validity? The Review of Higher Education, 35(1), 45–76.

Postareff, L., & Lindblom-Ylanne, S. (2008). Variation in instructors' description of teaching: Broadening the understanding of teaching in higher education. Learning and Instruction, 18, 109–120.

President's Council of Advisors on Science and Technology (2012). Report to the president. Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics. Washington, DC: Executive Office of the President.

Rogers, E. M. (1995). Diffusion of innovations (4th ed.). New York: Simon & Schuster.

Saroyan, A., & Snell, L. S. (1997). Variations in lecturing styles. Higher Education, 33(1), 85–104.

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Schoenfeld, A. H. (1999). Models of the teaching process. The Journal of Mathematical Behavior, 18(3), 243–261.

Schonwetter, D. (1993). Attributes of effective lecturing in the college classroom. The Canadian Journal of Higher Education 23(2), 1–18.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. Cognition and Instruction, 16(4), 475–522.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. Educational Researcher, 31(7), 15–21.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. CBE-Life Sciences Education, 12(4), 618–627.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. Review of Educational Research, 72(3), 387–431.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. Educational Researcher, 30(3) 23–28.

Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementation of peer instruction. Physical Review Special Topics–Physics Education Research, 5(2), 020101.

Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: Integrating active learning into a large lecture course. CBE-Life Sciences Education, 7(4), 361–367.

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., et al. (2011). Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program (UTeach Technical Report 2011–01). Austin: UTeach Natural Sciences, University of Texas at Austin.

Wertsch, J. V. (1991). Voices of the mind: A sociocultural approach to mediated action. Cambridge, MA: Harvard University Press.

West, E. A., Paul, C. A., Webb, D., & Potter, W. H. (2013). Variation of instructor-student interactions in an introductory interactive physics course. Physical Review Special Topics-Physics Education Research, 9(1), 010109.

Wieman, C. E. (2014). Large-scale comparison of science teaching sends clear message. Proceedings of the National Academy of Science, 111(23), 8319–8320.

Yon, M., Burnap, C., & Kohut, G. (2002). Evidence of effective teaching: Perceptions of peer reviewers. College Teaching, 50(3), 104–110.

Zhang, Z. H., & Linn, M. C. (2013). Learning from chemical visualizations: Comparing generation and selection. International Journal of Science Education, 35(13), 2174–2197.

Science
Education

Do Inquiring Minds Have Positive Attitudes? The Science Education of Preservice Elementary Teachers

CATHERINE RIEGLE-CRUMB,[1] KARISMA MORTON,[1] CHELSEA MOORE,[2] ANTONIA CHIMONIDOU,[3] CYNTHIA LABRAKE,[4] SACHA KOPP[5]
[1]Department of Curriculum and Instruction, STEM Education, University of Texas at Austin; [2]Department of Psychology, University of Massachusetts, Amherst; [3]UTeach Primary, College of Natural Sciences, University of Texas at Austin; [4]Department of Chemistry, University of Texas at Austin; and [5]College of Arts and Sciences, SUNY Stonybrook

Received 26 June 2014; revised 27 February 2015; accepted 30 March 2015
DOI 10.1002/sce.21177
Published online 14 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Owing to their potential impact on students' cognitive and noncognitive outcomes, the negative attitudes toward science held by many elementary teachers are a critical issue that needs to be addressed. This study focuses on the science education of preservice elementary teachers with the goal of improving their attitudes *before* they begin their professional lives as classroom teachers. Specifically, this study builds on a small body of research to examine whether exposure to inquiry-based science content courses that actively involve students in the collaborative process of learning and discovery can promote a positive change in attitudes toward science across several different dimensions. To examine this issue, surveys and administrative data were collected from over 200 students enrolled in the *Hands on Science* (HoS) program for preservice teachers at the University of Texas at Austin, as well as more than 200 students in a comparison group enrolled in traditional lecture-style classes. Quantitative analyses reveal that after participating in HoS courses, preservice teachers significantly increased their scores on scales measuring confidence, enjoyment, anxiety, and perceptions of relevance, while those in the comparison group experienced a decline in favorable attitudes to science. These patterns offer empirical support for the attitudinal benefits of inquiry-based instruction and have implications for the future learning opportunities available to students at all education levels. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**819–836, 2015

Correspondence to: Dr. Catherine Riegle-Crumb; e-mail: riegle@austin.utexas.edu

INTRODUCTION

While the metaphor of science, technology, engineering, and mathematics (STEM) as a pipeline has been rightly criticized as too simplistic, it is nevertheless clear that students' early experiences in science classrooms shape their future achievement and interests (Xie & Shauman, 2003). With the goal of better understanding and ultimately improving elementary science education in the United States, researchers and policymakers have increased their attention toward teachers. A growing body of research now focuses on the science content knowledge of elementary science teachers, as the subject matter expertise that they possess has clear implications for what students learn (Diamond, Maerten-Rivera, Rohrer, & Lee, 2014; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Kanter & Konstantopoulus, 2010; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Compared to secondary teachers, elementary teachers are much more likely to be trained as generalists and consequently less likely to have extensive content knowledge, and this pattern appears particularly pronounced for science (Haefner & Zembal-Saul, 2004). Clearly there are continued concerns about the need for teacher training programs and professional development to focus on increasing subject matter expertise.

Yet content knowledge is a necessary but insufficient characteristic of a successful teacher. Teachers' attitudes about the content they teach is another critical factor that has implications for classroom learning, and importantly, negative attitudes can exist independent of content-area expertise (Beilock, Gunderson, Ramirez, & Levine, 2010; Tosun, 2000). While there is comparatively less research on elementary teachers' attitudes toward science than math (Bursal & Paznokas, 2006), there is nevertheless evidence that many elementary teachers are not favorably inclined toward science. Such negative attitudes on the part of teachers can impact their students' attitudes toward science and can inhibit students' learning (Beilock et al., 2010; Jarrett, 1999; Ramey-Gassert, Shroyer, & Staver, 1996), creating a vicious cycle that must be interrupted. However, effectively changing how teachers view science is a challenging task (Mulholland & Wallace, 1996; Palmer, 2002).

The goal of this study is to examine whether inquiry-based science content classes might function to help break this cycle by improving preservice teachers' attitudes at a critical juncture before they begin their professional lives as classroom teachers. While there is much research on the positive impact of inquiry on outcomes for K–12 students (Borman, Gamoran, & Bowdon, 2008; Diamond et al., 2014; National Research Council, 2012b), we build on a smaller body of qualitative research regarding the benefits of inquiry instruction in college for preservice elementary teachers (Mulholland & Wallace, 1996; Palmer, 2002). Specifically, we suggest that students can become empowered and enthusiastic about the domain of science through active involvement in the process of inquiry, defined as engaging in the pursuit of scientific questions via data collection, experimentation, exploration, and discussion (National Research Council, 2000).

To examine this issue, we collected data from the *hands-on science* (HoS) undergraduate program at the University of Texas at Austin (Ludwig et al., 2013) to determine whether exposure to these inquiry-based science content courses promoted a change in the science attitudes of a sample of over 200 preservice elementary teachers. In exploring preservice teachers' attitudes, we go beyond the typical singular focus of much research on self-efficacy to instead consider personal attitudes toward science across several dimensions (van Aalderen-Smeets, Walma van der Molen, & Asma, 2012). Additionally, to ensure the robustness of our results, our design utilizes a comparison group of noneducation and nonscience majors enrolled in more traditional lecture-based science courses; our analyses also account for students' social and academic background. Our study offers promising evidence that science content classes in college can be a positive vehicle for changing the

attitudes of future elementary teachers, and subsequently has potential implications for the opportunities to learn both cognitive and noncognitive skills that are offered to future generations of elementary science students.

LITERATURE REVIEW

Framework: Considering Attitudes Across Multiple Dimensions

When exploring attitudes toward science, it is critical to recognize multiple relevant dimensions. Based on a comprehensive review of prior research on teacher attitudes, and motivated by the lack of substantive clarity and empirical transparency of most prior research, van Aalderen-Smeets et al. (2012) recently advanced a new theoretical framework to provide a cohesive model that captures primary teachers' attitudes toward science. Specifically, they developed a tripartite model of primary teachers' attitudes toward science that distinguishes between three overarching dimensions, each of which is composed of different elements: (1) *perceived control* (which includes elements such as self-efficacy), (2) *affective states* (which includes enjoyment and anxiety), and (3) *cognitive beliefs* (such as perceived relevance). This framework is informed by earlier theoretical models of attitudes (Eagly & Chaiken, 1993), but departs from prior models by considering perceived control (e.g., self-efficacy) as a core dimension, and furthermore defining behavioral intentions as a consequence rather than a component of science attitudes. Their model also calls for researchers to make a clear distinction regarding whether the focus is on teachers' personal attitudes toward science or their professional attitudes toward teaching science, as studies that combine teachers' views of science as a domain with their views on science instruction in their own classroom into one empirical scale blur the object of teachers' attitudes, making substantive interpretation difficult.

The three dimensions of attitudes (whether personal or professional) advanced by van Aalderen-Smeets et al. (2012) are logically related to one another. For example, Bursal and Paznokas(2006) found that teachers with higher levels of self-efficacy had lower anxiety, a finding supported by several other studies (Bleicher, 2007; Palmer, 2002). Yet while related, they nevertheless capture somewhat distinct thoughts and beliefs. For instance, an individual might expect to master an activity or believe that it is useful, but nevertheless find it is unappealing (e.g., flossing their teeth) or even anxietyproducing (e.g., running 10 miles). Therefore, considering elements of all three dimensions is critical to developing a comprehensive picture of teachers' attitudes toward science.

Yet most of the literature about the attitudes of elementary science teachers (either preservice or in-service) focuses on their perceived control in the form of self-efficacy, as there is relatively scant research on either the cognitive beliefs or affective attitudes of teachers (van Aalderen-Smeets et al., 2012). Thus, a key contribution of our study is our consideration of elements of all three dimensions of attitude to more fully capture the complexity of preservice teachers' attitudes toward science.[1] Below we discuss the prior literature on the science attitudes of preservice elementary teachers in more detail, including how such attitudes may influence the outcomes of future students. Because our research questions and subsequent empirical analyses focus on preservice teachers well before they

[1] In this paper, we address all three dimensions of van Aalderen-Smeets et al.'s (2012) theoretical model, but do not discuss (or model) every element within each dimension. For example, while the authors include context dependency as an element that falls under the dimension of perceived control, we do not address this here as it refers to the support that practicing teachers receive from their administrators and therefore is not particularly relevant for a study concerning the personal (rather than professional) attitudes of preservice teachers.

actually enter the elementary classroom, we concentrate on literature on personal attitudes toward science. We then turn to a discussion of why inquiry-based science content classes for preservice teachers have the potential to increase individuals' attitudes across all three dimensions.

Perceived Control: Considering Self-Efficacy. A key element of the attitudinal dimension of perceived control is self-efficacy, which according to Bandura's (1977, 1982) foundational work is defined as an individual's belief that she can successfully master a situation or deal with an obstacle that arises. An individual's self-efficacy has logical implications for her subsequent behaviors and choices, as she is likely to attempt to avoid those situations or activities where she does not feel she can be efficacious, and persist where she feels confident that she can be successful. While research has demonstrated that in-service elementary teachers exhibit low levels of efficacy in their science teaching (see, for example,Atwater, Gardner, & Kight, 1991; Harlen, 1997; Ramey-Gassert et al., 1996), not surprisingly this pattern is also evident among preservice teachers, who feel less efficacious about their own ability to learn science. For example, Skamp (1991) determined that less than half of the preservice elementary teachers in his study reported having even a fair amount of confidence in their science ability. Similarly, Bleicher's (2007) study of preservice elementary teachers found that participants in a science methods class exhibited low scores on science self-efficacy scales. Low efficacy is often attributed to prior negative educational experiences in science. For example, in a study of five preservice teachers, Mulholland and Wallace (1996) found that their respondents reported very little confidence in their own science abilities and attributed this to negative experiences in their own science schooling as children.

Affective States: Considering Enjoyment and Anxiety. Individuals' affect toward a domain represents another attitudinal dimension. As articulated by van Aalderen-Smeets et al. (2012), the dimension of affect can be further categorized into the positive element of enjoyment and the negative element of anxiety. Beginning with the former, Liang and Gabel (2005) found that most preservice elementary teachers in their study reported that science had never been enjoyable for them. These teachers also indicated that they took science courses only because it was required for their degree program. Smith (2000) and Howes (2002) found similar results. Additionally, the preservice participants in all of these studies attributed their low levels of enjoyment to negative experiences in either (or both) their high school and college science courses, a point to which we will return to later.

Anxiety captures the negative aspect of the affective dimension. Early research by Mallow (1981; also Mallow & Greenburg, 1983) as well as Westerback (1984) define science anxiety as the fear, worry, or apprehension that some individuals experience when presented with the task of learning science. While research on the math anxiety of preservice teachers is extensive (Udo, Ramsey, & Mallow, 2004), and research considering both math anxiety and science anxiety find a strong association between the two (Cady & Rearden, 2007), there is comparatively little research focusing specifically on science anxiety (Bursal & Paznokas, 2006). Yet several studies do provide evidence that preservice elementary teachers report relatively high levels of science anxiety about teaching science, as well as more generalized anxiety about learning science themselves (Cady & Rearden, 2007; Udo et al., 2004; Westerback & Long, 1990).

Cognitive Beliefs: Considering the Relevance of Science. Finally, we discuss research on preservice elementary teachers' beliefs in the relevance of science, a critical element

of the attitudinal dimension of cognitive beliefs. The limited extant research on this topic focuses on perceptions of science as useful or relevant for society and for them personally, and finds that preservice elementary teachers report generally favorable views. Specifically, in a study of 200 preservice teachers, Coulson (1992) used a survey instrument that included a personal usefulness science scale and found that on average, respondents indicated moderate levels of agreement. Cobern and Loving's (2002) study at a large Midwestern university also found that on Likert scales measuring the perceived importance of science to all citizens, preservice elementary teachers on average agreed with this sentiment. These findings echo the sentiments of the general population, which generally regard science and technology as useful for making their lives better (Evans & Durant, 1995; Kohut, Keeter, Doherty, & Dimock, 2009). Therefore among the three attitudinal dimensions discussed, promoting preservice teachers' views of the relevance of science is perhaps somewhat less of a pressing problem than promoting both their perceived control in the form of science efficacy and their affective attitudes of enjoyment and anxiety.

Examining the Impact of Teacher Attitudes

There is a logical connection between teachers' attitudes toward a subject and student outcomes, both in terms of impacting students' opportunities to learn science and their own developing attitudes. First, research indicates that the negative attitudes toward science held by many elementary teachers are likely to result in less coverage of science content and less engaging and effective instruction. Teachers who are not confident in their own science knowledge are likely to worry that they cannot effectively answer students' questions nor keep them engaged with interesting activities (Jarrett, 1999). Consequently, elementary teachers who feel less efficacious and more anxious are likely to try to avoid teaching science and spend less time teaching when they cannot avoid it altogether (Brownlow, Jacobi, & Rogers, 2000; Pine et al., 2006; Ramey-Gassert et al., 1996). Appleton and Kindt (1999) also found evidence that avoidance occurred when teachers did not find science to be as relevant as other subjects such as English or math. One such teacher reported that "If you're running out of time in the week, you think 'Oh I just won't worry about that science activity'" (p. 162).

When teachers do teach science, their negative attitudes can impact their pedagogical practices and subsequently their capacity to reach and engage students. Appleton and Kindt's (1999) study of in-service elementary teachers found that those exhibiting low confidence were less likely to engage their students in hands-on learning. Ramey-Gassert et al. (1996) found that elementary teachers with low science self-efficacy had a minimal desire to engage in professional development activities that could improve their teaching of science. Lack of belief in the usefulness or relevance of science could also logically result in a reduced effort to provide engaging instruction to students.

Additionally, the negative attitudes toward science held by many elementary teachers can result in the socialization of students toward a negative stance toward science. As students look to their teachers as role models and authorities, they begin to mimic and potentially internalize such attitudes as their own (Jussim & Eccles, 1992; McKown & Weinstein, 2002). This can lead to a vicious cycle where negative attitudes such as anxiety and low efficacy are passed from teachers to their students, and therefore from one generation to the next. For example, Beilock et al. (2010) found that elementary teachers' math anxiety influenced students' own attitudes toward math, with girls in particular more likely to evidence declining math attitudes and increasingly gender-stereotyped views of math over the course of the year. This particular study highlights additional concerns regarding gender rolemodeling; as most elementary teachers are female, and a substantial number

of them exhibit negative attitudes toward science (as well as math), this could be a strong conduit through which young girls begin to believe that these subjects are less interesting, less important, and overall less-suited for them. Furthermore, as students' own attitudes decline, so does their engagement and motivation to learn, which further impacts their achievement (Eccles, 1994).

In sum, the research literature provides compelling evidence that teachers' negative attitudes can impact students' learning and attitudinal outcomes, and as such are a critical issue that needs to be addressed. To effectively break this cycle, necessitates intervening to change teachers' attitudes before they enter the classroom. In short, the education of preservice teachers is an ideal place to focus.

Educating Preservice Teachers: Inquiry as a Tool for Improving Attitudes

As mentioned earlier, the negative attitudes of preservice elementary teachers can at least in part be traced back to their own negative experiences with science classes in high school and in college. For example, Liang and Gabel (2005) found that preservice teachers attributed their lack of enjoyment of science to their prior experiences in classes dominated by lectures that necessitated copious amounts of note-taking and memorization. Similarly, Smith (2000) reports that preservice elementary teachers in his study often complained about the boredom of their procedure-based high school and college science courses. Simply put, many preservice teachers have had little exposure to inquiry-based science instruction in their lives as students. We posit that exposure while in college to pedagogy that actively involves them in the process of scientific discovery can ultimately help them to see that science is meaningful, interesting, and accessible, thereby changing their attitudes toward science.

Several studies support the supposition that inquiry classes can promote such changes, in spite of the possibility of some initial student frustration with an approach that deviates from teachers' didactic presentations of the "right" answer (Volkmann, Abell, & Zgagacz, 2005). For example, in a qualitative study of preservice teachers at a large university in the southwest, Kelly (2000) found that after completing an active, inquiry-based science methods course, most participants reported that their interest in science had increased. Kelly (2000) attributes this shift to the use of hands-on explorations and discussions where students came to embody the process of scientific inquiry by formulating and exploring ideas. Similarly, in a study of preservice elementary teachers, Palmer (2002) interviewed four participants who reported that their attitudes toward science had changed from negative to positive due to the excitement of inquiry-based lessons. A study of five teachers by Mulholland and Wallace (1996) reported similar results. Finally, in a quantitative study of 112 preservice elementary teachers, Jarrett (1999) found that an inquiry-based science methods class increased the participants' personal interest in science. The author attributed this change to preservice teachers becoming active agents in the classroom and learning to view science as a process of discovery. Thus, a small body of research finds that inquiry-based pedagogy in science educational methods courses can have a positive impact on preservice elementary teachers' attitudes.

CURRENT STUDY

Building on the insights of this prior research, we posit that *required science content courses* could also represent a powerful venue for change for college students at the beginning stages of preparing for their careers as future teachers. Specifically, the purpose

of this study is to examine whether inquiry-based science content courses promote a change in attitudes toward science among preservice elementary teachers. The courses are part of the HoS undergraduate program, developed at The University of Texas at Austin in the College of Natural Sciences, with cooperation from the College of Education. HoS is required of all students in the elementary education program and covers four semesters of science courses with a curriculum that is composed heavily of the topics that preservice teachers will be expected to teach their students once they become teachers (Ludwig et al., 2013). The design of the HoS program is based on the Physics and Everyday Thinking (Goldberg, 2008) framework that centers around the development of students' physical science understandings through experimentation and follows the example of work from Western Washington University extending this framework to other disciplines (Nelson, 2008). The curriculum is based upon big ideas in science, with specific emphasis on the themes of Matter and Energy, which are integrated across different science disciplines. The course sequence focuses on physics in Semester 1, chemistry and geology in Semester 2, biological systems in Semester 3, and astronomy and earth science in Semester 4.

HoS classes were designed to utilize the essential elements of inquiry-based learning as defined by the influential National Research Council report on inquiry (Forbes, 2011; National Research Council, 2000); specifically, students engage in scientifically oriented questions by collecting, organizing, and analyzing data. From that data they formulate explanations, connect them to scientific knowledge, and subsequently evaluate their explanations in contrast to alternative explanations. Additionally, students share and justify those explanations with others.

The HoS program is best categorized as teacher-directed or guided inquiry (Cuevas, Lee, Hart, & Deaktor, 2005; National Research Council, 2000; Volkmann et al., 2005). Initial questions are posed by the instructor and all activities are carefully designed to present opportunities for students to confront misconceptions, with both topics and skills developed in a structured progression. The instructor does not lecture but probes student knowledge and offers helpful nudges in the right direction, working with students both in small groups and via whole class discussions to construct knowledge and develop explanations (Volkmann et al., 2005). Finally, consistent with inquiry approaches to learning, the course emphasizes big ideas in science while engaging students in hands-on learning and constant exploration and discussion while in groups with their peers (National Research Council, 2000, 2012b).

An example of a typical 2-hour class session is a lesson on sound that begins with the following teacher-generated questions: *How does sound travel through a room? How does sound travel from your classmate to you?* Students then work in groups of three or four to answer these questions and discuss their initial ideas or preconceptions; they subsequently share their ideas with the entire class, so that there is a collective knowledge of alternative ways of thinking about how sound travels. Working in their small groups, students then engage in a series of experiments, using an "airzooka" and candles, a string telephone, and a loudspeaker, that require that they gather data and then generate models based on these data. They are regularly prompted by the instructor to make predictions and connect trends to other previously seen concepts. Finally, students reflect and summarize key ideas and connect them to other contexts by using evidence-based reasoning from the data collected in their experiments. These questions are the basis for whole-class discussions, where students present and justify their ideas to their peers. Students typically end lessons by writing a scientific narrative summarizing how their thinking was revised from their initial ideas, drawing on evidence gained through the experiments and whole class sharing and discussion. For example, a student might share how her initial idea that "sound travels as a wave" has become more sophisticated, discussing the role of vibrations and the transfer of mechanical energy from the source to the receiver.

Our examination of whether the inquiry-based science content courses of the HoS program promote a change in attitudes toward science among preservice elementary teachers is described in detail below. Here, we briefly note that our study contributes to the literature a rigorous quantitative examination of change in attitudes across multiple dimensions among preservice elementary teachers, utilizing a comparison group of students in more traditional science content courses, as well as accounting for differences among students in academic and social background characteristics that may have implications for their attitudes.

Data and Method

Analytic Sample. Our analytic sample includes 238 preservice elementary teachers who enrolled in the HoS program between Fall 2010 and Spring 2012. Students completed presurveys prior to the start of the first course and postsurveys at the end of the second course in the sequence.[2] Surveys were done online, and were administered during class by members of the research team. Additionally, our sample includes a comparison group composed of 263 nonscience and noneducation majors enrolled in traditional lecture-style undergraduate science courses in Spring 2012. While our design falls shorts of a pure treatment versus control comparison, we chose courses that represent what our sample of preservice elementary teachers would have been required to take in the absence of the HoS program. Therefore, our comparison group includes students who were enrolled in either an introductory-level chemistry or biology class. These students were surveyed at the beginning and end of the semester. As the time period between pre- and postsurveys is longer for HoS students than for non-HoS students, we discuss the implications and limitations of this comparison later.

Administrative records were linked to student surveys, allowing us to examine the demographic and academic background information of students in our sample and to consider how HoS students differed from those in the comparison group. Not surprisingly given the gender composition of the elementary teacher population nationwide, HoS students were overwhelmingly female (95%), compared to 65% of our comparison sample of noneducation and nonscience majors. There were no statistically significant differences between the two groups of students by mother's education; for race/ethnicity, there were significantly fewer students who identified as American Indian among HoS students compared to non-HoS students. Regarding academic background prior to college entry, HoS students had lower SAT math scores than their fellow students taking typical science classes by more than one-third of a standard deviation (see Table 1). To ensure that any differences observed between the two groups in their changes in attitudes over time are not confounded by differences in their background characteristics, our subsequent multivariate models will control for all of these factors.[3]

Student Attitude Surveys. The pre- and postsurveys administered to both groups consisted of 21 items geared toward assessing student attitudes toward science. To examine multiple dimensions of attitudes, we selected items from preexisting surveys to measure

[2]At the time of this study, students were only required to take the first two semesters of the total four semester sequence.

[3]Mother's education level is an ordinal variable with the following categories: (1) did not attend high school, (2) attended high school but did not graduate, (3) high school diploma or GED, (4) some college, (5) earned associate's degree, (6) bachelor's degree, (7) graduate or professional degree. Math SAT score and mother's education level were imputed for those students who had missing values. We utilized information on gender, race, high school rank, family income, and father's education to conduct single imputation using Stata. Analyses using list-wise deletion produced similar results.

TABLE 1
Descriptive Statistics

	Hands-on Science (HoS) Students		Non-*HoS* Students	
	$N = 238$		$N = 263$	
Gender				
Female***	94.5%		65.4%	
Race/ethnicity				
White (non-Hispanic)	63.4%		60.1%	
Black	2.9%		3.4%	
Hispanic	21%		19%	
Asian	11.3%		14.4%	
American Indian**	0.4%		2.7%	
Native Hawaiian/Pacific Islander	0.8%		0.4%	
Other background characteristics				
	Mean	SD	Mean	SD
SAT math score***	574.87	75.60	607.62	73.39
Mother's educational level	5.10	1.60	5.07	1.69

***$p < .001$, **$p < .01$, *$p < .05$, $p < .10$.

confidence (an element of perceived control)[4], enjoyment and anxiety (positive and negative elements of affective states), and relevance (a key element of cognitive beliefs). To measure anxiety, we used the Math Anxiety Rating Scale (Hopko, 2003) and substituted the word science for all references to math. To measure all other attitudes, we selected items developed by the National Center for Education Statistics (www.nces.ed.gov), and used in national surveys, including the Educational Longitudinal Study, the High School Longitudinal Study, and the U.S. component of the Trends in International Mathematics and Science Study. Principal component analyses with promax rotation using the complete set of 21 survey items revealed four factors with Eigenvalues greater than one. We subsequently created separate scales (described below) composed of the individual items that loaded onto each of the four factors. The scales clearly corresponded to the elements of confidence, enjoyment, anxiety, and relevance; using the full sample, the Cronbach's alpha for each scale was .8 or higher.[5]

The *confidence* scale is composed of four items that gauge students' level of confidence when engaging in scientific activities. The items were as follows: "I have always done well in science," "Science is not one of my strengths" (reverse coded), "I am confident that I can understand the most difficult material presented in my science textbooks," and "Science

[4] Because the questions ask students to reflect more on their assessment of their current success in science, we refer to this scale as measuring confidence rather than self-efficacy for future success. However, prior research has consistently noted a very strong correspondence between indicators of self-confidence and self-efficacy (Wigfield & Eccles, 2000).

[5] To further test the reliability of the four scales, we calculated Cronbach's alphas separately for (a) treatment and reference groups; (b) first year cohort and second year cohorts (for treatment group), and (c) survey responses at time 1 and time 2 (for different cohorts and for different treatment groups). In all, we calculated alpha reliabilities for each of our four scales for 15 different subsamples with remarkably consistent results ranging from a low of .77 to a high of .88. (exploratory factor analyses using each of these groups also yielded the same four factor model).

has always been one of my best subjects." The five items in the *enjoyment* scale capture students' level of positive affect for science and included "I enjoy learning science," "I look forward to going to science courses," "Science is fun," "I like science," and "Science is boring" (reversecoded for inclusion in the scale). Categories of response were the same as those for the confidence scale.

For each of the eight items in the *anxiety* scale, students were asked to indicate how much the situation made them feel anxious or worried. Similar to the other measures, the responses ranged from (1) not at all anxious to (5) very much anxious or worried. The items included "Looking through the pages in a science text," "Thinking about an upcoming science test one day before the test," "Reading and interpreting a scientific graph, chart, or illustration," "Taking an exam in a science course," "Watching and listening to a teacher explaining a scientific concept or phenomena," "Waiting to get a science test returned in which you expected to do well," "Walking on campus and thinking about a science course," and "Being given a 'pop' quiz in science class."

Finally, the four items making up the *relevance* scale include questions that focused on students' perception of the meaning or usefulness of science in their daily lives. The items are as follows: "The subject of science is not very relevant to most people," "It is not important for most people to understand science," "I think learning science will help me in my daily life," and "Science is important to me personally." Students indicated the extent to which they agreed or disagreed with the statements with possible responses ranging from strongly agree (1) to strongly disagree (5).

ANALYSES AND RESULTS

To examine whether inquiry-based science content courses promote a positive change in attitudes toward science for preservice elementary teachers, we begin by comparing the means on the pre- and postmeasures of our four attitudinal scales for HoS students. Results of paired t-tests indicate a statistically significant improvement over time for each scale. Specifically, for confidence, the mean increased from 2.61 to 2.98 ($p < .001$), reflecting a 0.37 point increase or almost half of a standard deviation change from pre to post. HoS students' science enjoyment increased by about a fourth of point (and about a third of a standard deviation), from the presurvey mean of 3.24 to a postsurvey mean of 3.51 ($p < .001$). The largest change was observed for the anxiety scale, where the average decreased (meaning that students became less anxious over time) from a presurvey mean of 3.11 to a postsurvey mean of 2.63 ($p < .001$), a difference of almost half of a point and approximately two-thirds of a standard deviation. Finally, as discussed earlier, students view science as a relevant domain, as evidenced by a relatively high presurvey mean of 3.66 on the utility scale. Nevertheless, they slightly increased their views of the relevance of science after taking HoS courses ($p < .05$) to a postsurvey mean of 3.74, a change of about a tenth of a standard deviation.

Subsequently, we turn to an examination of how the changes we observe for HoS students compare to changes in attitudes toward science among students taking more traditional lecture-based science content courses. Here, we utilize multilevel mixed-effects models, an extension of regression analysis that is similar to a mixed-design analysis of variance and appropriate when data are nested. For this study, repeated measures of attitudes are nested within individuals with time treated as a random effect (Rabe-Hesketh & Skrondal, 2008).The goal of this analysis was to determine whether the change in different dimensions of science attitudes observed for HoS students was similar or different than that observed for non-HoS students while controlling for students' background characteristics (which

TABLE 2
Regression Analyses Predicting Attitudes to Science[a]

	Model 1 Confidence	Model 2 Enjoyment	Model 3 Anxiety	Model 4 Relevance
Hands-on science (HoS) students (ref = non-HoS students)	−0.391***	−0.163*	0.106	−0.051
	(0.071)	(0.072)	(0.066)	(0.059)
Time (change from pre- to postsurvey)	−0.127**	−0.131***	0.074~	−0.116**
	(0.039)	(0.039)	(0.041)	(0.036)
Time × *HoS*	0.500***	0.391***	−0.553***	0.200***
	(0.056)	(0.056)	(0.060)	(0.052)
Female	−0.402***	−0.395***	0.311***	0.046
	(0.085)	(0.084)	(0.076)	(0.069)
Race/ethnicity (ref = white)				
Black	0.078	−0.069	0.018	−0.159
	(0.188)	(0.185)	(0.168)	(0.149)
Hispanic	−0.136	0.054	0.055	−0.023
	(0.093)	(0.091)	(0.083)	(0.073)
Asian	−0.314**	−0.110	0.232**	−0.002
	(0.098)	(0.097)	(0.088)	(0.078)
American Indian/Alaska native	−0.591*	−0.181	0.287	0.112
	(0.251)	(0.249)	(0.224)	(0.204)
Native Hawaiian/other Pacific Islander	0.635	0.380	−0.524	0.537~
	(0.418)	(0.407)	(0.374)	(0.322)
SAT math score	0.002***	−0.000	−0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mother's education level	−0.038~	−0.013	0.029	−0.003
	(0.022)	(0.022)	(0.020)	(0.017)
Constant	2.680***	4.179***	3.748***	3.762***
	(0.320)	(0.315)	(0.288)	(0.255)

[a]Coefficients calculated from a multilevel regression model in Stata where observations of attitudes are nested within individuals.
***$p < .001$, **$p < .01$, *$p < .05$, ~$p < .1$; standard errors in parentheses; $n = 501$.

is particularly important as our two groups of students differed by gender and math SAT scores, both factors which likely predict attitudes).

Table 2 displays the results of separate analyses for each dependent variable. The first row displays the coefficient comparing HoS and non-HoS students on the presurvey (or time 1) for each attitudinal dimension. The second row displays the average change over time between the pre- and postsurvey, while the third row displays the interaction between student group (HoS or non-HoS) and time. The change in attitudes from pre- to postsurvey for HoS students is calculated as the sum of the main effect of time and the interaction term, while change for non-HoS students is captured by the main effect of time only. To simplify the presentation of results, we include figures for each attitudinal outcome (Figures 1–4) that display the changes over time for each group, adjusted for the social and academic characteristics discussed above.[6]

[6]Figures 1–4 display the adjusted pre and post means for each group. These are calculated using a postestimation command in Stata where all other variables in the model(other than student group, time, and the interaction) are set to the mean (or alternatively to the mode for categorical variables).

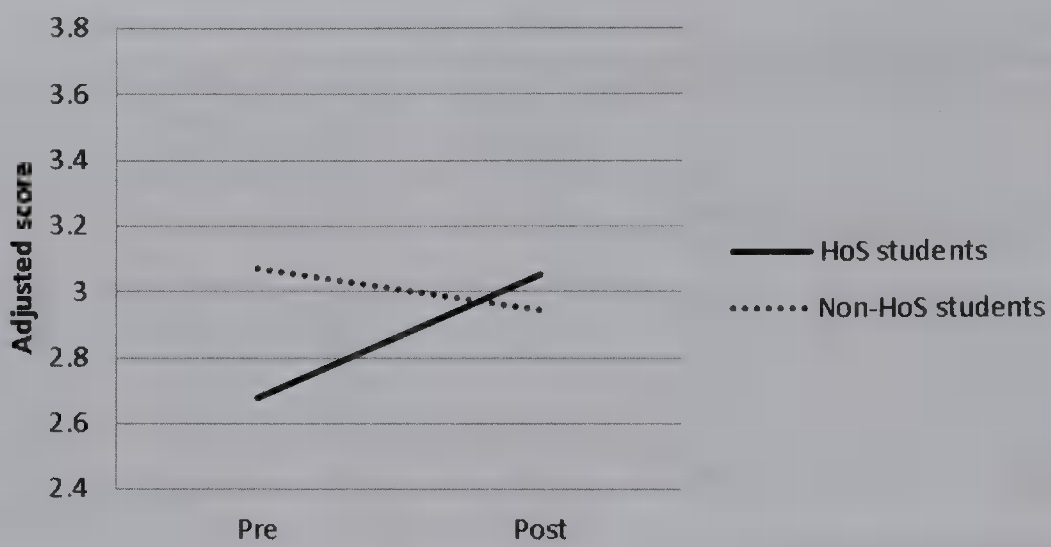


Figure 1. Confidence.

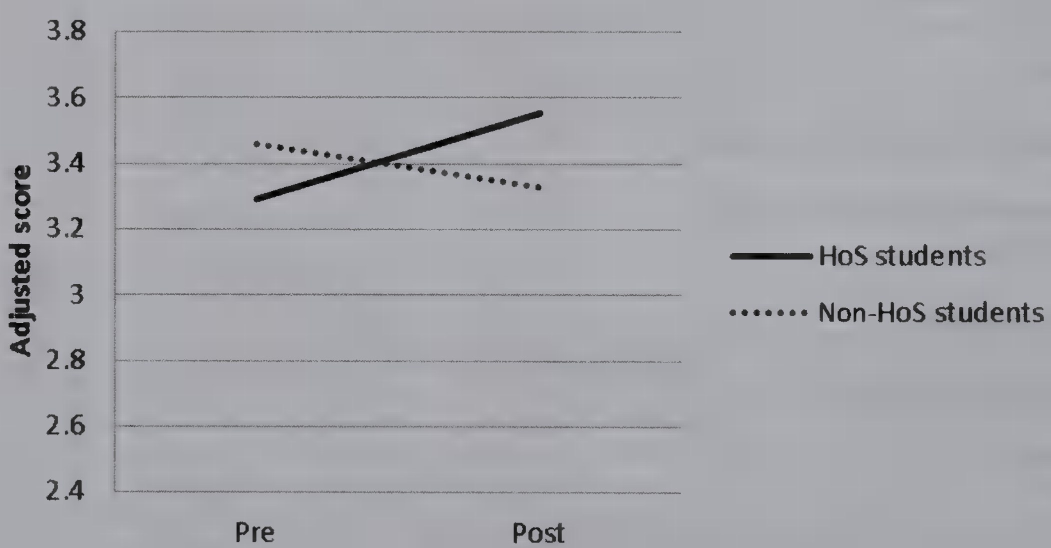


Figure 2. Enjoyment.

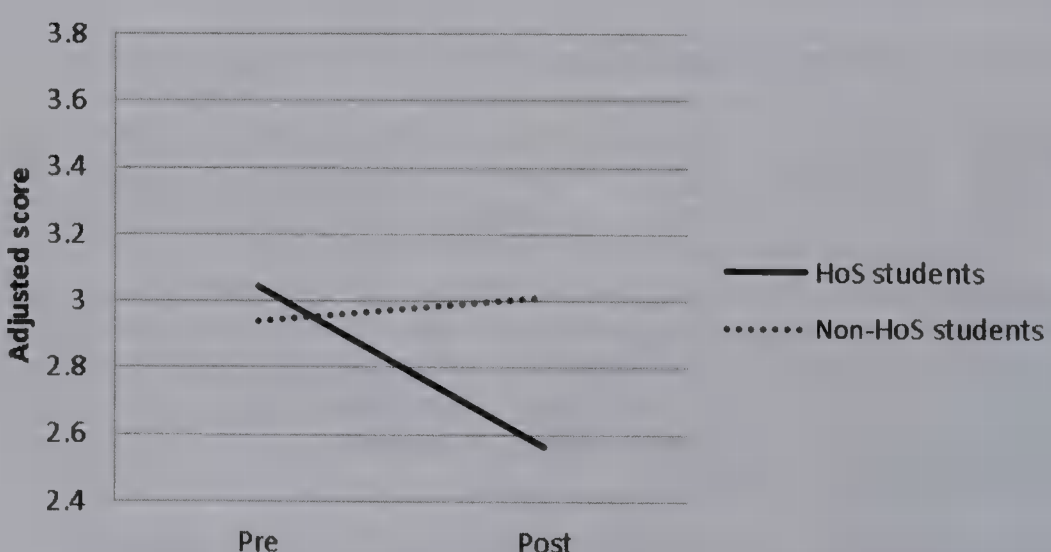


Figure 3. Anxiety.

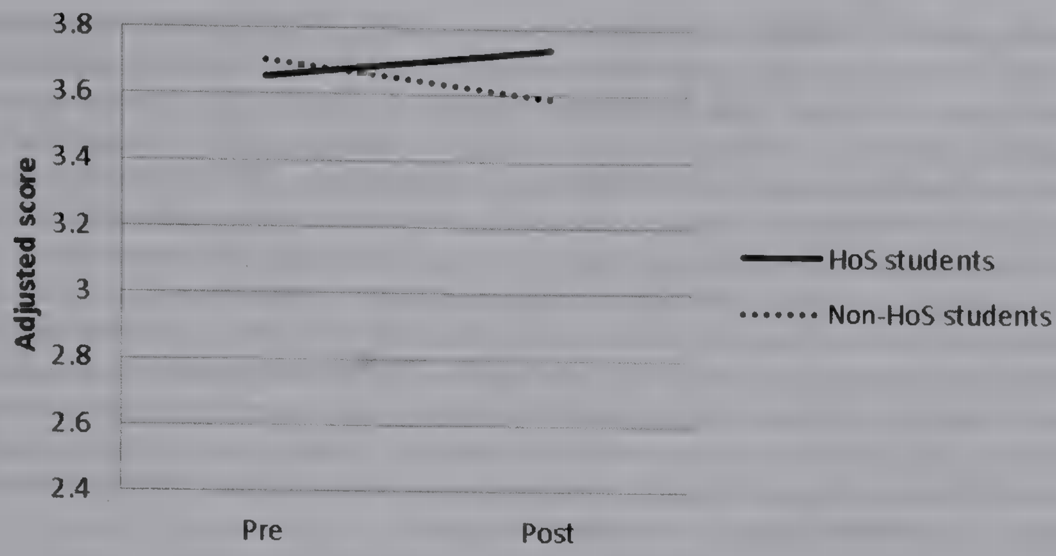


Figure 4. Relevance.

Beginning with the first column predicting *confidence,* the results reveal that compared to their peers in traditional science classes, HoS students reported significantly lower science confidence on the presurvey (–.391***). The coefficient for time is negative and significant, yet the interaction between time and student group is positive and significant, indicating that HoS students increased their confidence over time relative to non-HoS students. To help clarify the patterns for the two groups, Figure 1 displays the trends for each group. Here, we see clearly that changes in attitudes occurred for both groups in opposite directions. While HoS students initially had lower confidence than their non-HoS peers, they significantly increased their confidence over time. In contrast, non-HoS students significantly decreased their science confidence (–.127***), such that their confidence was slightly lower than non-HoS students on the postsurvey.

Returning to Table 2, we see a similar pattern when predicting change in science *enjoyment.* HoS students initially reported significantly lower levels of enjoyment than their peers in more traditional classes (–.163**). However, once again we see a negative main effect of time but a positive and significant interaction between student group and time, indicating opposite directions of change for each group. As displayed graphically in Figure 2, HoS students significantly increase their enjoyment over time, while on average non-HoS students report a decrease in their affect toward science (–.131***), and end their course reporting lower enjoyment than HoS students.

Regarding changes in *anxiety*, we note that HoS and non-HoS students do not differ significantly on the presurvey. The main effect of time is positive and borderline significant, yet the interaction term reveals a statistically significant difference between the two groups' average change in anxiety. Figure 3 clearly shows the marked decrease in anxiety from the pre- to the postsurvey for HoS students. For their non-HoS peers, however, the figure shows a slight increase in anxiety (~.074).

Finally, Table 2 displays the results for models predicting changes in attitudes toward the relevance of science. HoS and non-HoS students do not differ significantly on the presurvey. But once again the interaction term reveals a statistically significant difference between groups in change over time, and the sign of the coefficient is positive in contrast to negative main effect of time. Figure 4 displays the disordinal patterns for the groups. HoS students' views of relevance increase a small amount from the pre- to the postsurvey, whereas their peers' perceptions of the relevance of science decreases (–.116**).

Finally, while the main focus of our analyses was to assess differences between HoS and non-HoS students regarding changes in their science attitudes, our multivariate analyses revealed patterns consistent with prior research, namely that females are significantly less confident in their science ability and report significantly less enjoyment and more science anxiety than their male peers (Correll, 2001; Eccles, 1994). Our models also indicate some evidence of racial/ethnic differences in attitudes (see the models predicting confidence and anxiety), as well as differences by prior math achievement, as those with higher scores on the math portion of the SAT report significantly higher levels of confidence in their science ability and less anxiety. Given the differential distribution of HoS and non-HoS students on several of these covariates (see Table 1), including them in our models generally decreased the size of the difference between groups on the presurvey (e.g., the more male composition of the non-HoS group helped to account for their initially stronger math confidence), and also revealed larger differences between the groups on the postsurvey than could be detected in simpler models that did not account for these factors.

DISCUSSION AND CONCLUSION

The goal of this study was to focus on future elementary teachers' views toward science at a point when they are still explicitly occupying the role of a learner, before they are tasked with assuming the role of a science teacher. Specifically, we investigated whether enrollment in HoS, a program of inquiry-based science content courses, promoted a favorable change in the science attitudes of a sample of over 200 preservice elementary teachers. Building on the theoretical framework advanced by van Aalderen-Smeets and colleagues (2012), our study examined changes in attitudes on multiple dimensions, and also utilized a comparison group of noneducation/nonscience majors to help contextualize the changes observed for our focal sample of elementary preservice teachers.

Data analyses reveal a remarkably consistent and positive story for HoS students; students significantly changed their views toward science from the pre- to the postsurvey, such that after participating in inquiry-based content courses they reported more confidence in their skills as science learners, more enjoyment and less anxiety toward science, and perceived it as more relevant. Conversely, patterns for those in the comparison group revealed a decline in favorable attitudes toward science after enrolling in a traditional, lecture-based content course. Importantly, these results are independent of differences between HoS and non-HoS students (e.g., gender and SAT math score) that are associated with attitudes; therefore, our analyses indicates that it was differences in the courses that the two groups of students enrolled in, rather than characteristics of the individual students themselves, that led to changes in attitudes.

Our study has several likely implications for the future classrooms of preservice teachers, as prior research suggests that teachers with negative views toward science may both socialize their young students to develop similar views, as well as offer less science instruction in class due to a desire to avoid the subject (Bursal & Paznokas, 2006; Jarrett, 1999). Thus, to the extent that HoS students have more positive attitudes toward science as a result of inquiry-based instruction, we have positively intervened to disrupt the vicious cycle of elementary school teachers passing their negative views of science onto the next generation. Instead, future elementary students could ultimately be the beneficiaries, having teachers who are favorably inclined and excited about teaching science and therefore spend more time and focus on it.

Additionally, we suggest that our results have potential implications for gender equity in the classroom, particularly because, while all observed changes for HoS students were in a favorable direction, the largest changes were for decreasing anxiety. Recent research by

Beilock et al. (2010) offered evidence that the math anxiety of female elementary teachers had a negative impact on their female students in particular. The authors argue that due to the inclination for children to more strongly connect to adults of the same gender as role models, young girls in the classroom were more susceptible to teachers' math anxiety, and consequently exhibited more negative attitudes of their own as well as lower math achievement. Their study provides powerful evidence that teacher role-modeling can be a key factor that leads to the development of the gender gap in math as early as elementary school (Beilock et al., 2010). Such a pattern can be logically extended to science anxiety; thus by intervening to decrease the anxiety of (predominantly female) preservice elementary teachers, more young girls could have the opportunity to interact with a positive female role model, thereby thwarting emerging gender disparities in children's views and performance (Eccles, 1994).

Our study also has potential ramifications for the type of instruction and classrooms experienced by future elementary students. As research suggests that preservice teachers tend to teach their students in ways similar to how they were taught (Gess-Newsome & Lederman, 1999), programs such as HoS can serve as a powerful model for implementing inquiry in their own classrooms, and thus ultimately contribute to greater uptake of recommended science reform. The Next Generation Science Standards call for elementary teachers to engage their students in inquiry-based practices, as well as emphasize disciplinary core ideas in the classroom (National Research Council, 2012b). The HoS classes offer preservice teachers the critical opportunity to develop an understanding of and real experience with inquiry-based teaching and learning that is consistent with these standards. We concur with Volkmann et al. (2005), who argue that "if learning through inquiry is to become a reality in today's schools, then university science courses must model inquiry so that pre-service teachers may experience it" (p. 867). Programs such as HoS have the power to do exactly this, and thereby help break the cycle of teacher-centered didactic instruction.

While the primary focus of this study is the educational experiences of preservice teachers during college and the subsequent implications for future elementary classrooms, our study also speaks to the need to improve undergraduate science education more broadly. A recent meta-analysis of student academic performance in STEM undergraduate courses provides evidence that traditional lecture formats lead to higher failure rates and lower achievement when compared to classes that are more constructivist based (Freeman et al., 2014). Our study focuses on attitudinal rather than performance outcomes, and in doing so heeds recent calls by the National Research Council to examine a more comprehensive range of student outcomes at the postsecondary level (National Research Council, 2012). Specifically, we find that, even after adjusting for differences in social and academic background between our two groups (preservice education students and noneducation/nonscience students), enrollment in traditional lecture classes has the opposite effect of enrolling in inquiry-based content classes. We suggest that the decline in confidence, affect, and utility, as well as a slight increase in anxiety that we observed for students in traditional lecture-based science content classes, are important consequences of their relatively low engagement in the classroom, and perhaps linked to patterns of lower performance documented elsewhere (Freeman et al., 2014; Seymour & Hewitt, 1997).

As with any study, ours has limitations. First, we note a lack of parallel time frames for our focal HoS students (two semesters between pre- and postsurveys) and our comparison group (one semester between pre- and postsurveys). Therefore, it is possible that part of the positive change in attitudes observed for HoS students is due to the longer exposure period; while we cannot dismiss this possibility entirely we are nevertheless skeptical that a shorter survey window for HoS students would have substantively changed our findings regarding opposite directions of change for the two groups. Indeed, we did collect postsurveys for a

small subsample of HoS students at the end of the first semester, and the results, although slightly weaker in magnitude, were statistically significant and in the same positive direction as the full sample of HoS students included here.

Additionally, we did not collect data to assess what features of these inquiry-based classes students found most favorable; for instance, it could be that the significant amount of time spent on group work was a particularly influential factor leading to their change in attitudes (Park Rogers & Abell, 2008). We think this is an important area for future research to address, to better ascertain which aspects of inquiry-based classrooms are most effective at promoting favorable shifts on different dimensions of science attitudes. More long-term studies are also needed to assess whether and how the potential implications we discuss above come to fruition and make a difference for elementary students in the science classroom.

Finally, it is important to point out that designing and implementing inquiry-based science content courses that depart from the typical lecture-based format of most postsecondary instruction is certainly not without its challenges and difficulties (Allen & Tanner, 2005; Armbruster, Johnson, & Weiss, 2009). One such obstacle is convincing science instructors and university administrators of the benefits of inquiry-based instruction for their students. Our study contributes to the small number of studies that offer robust empirical evidence on this topic (Seymour, 2002), and in doing so offers additional support to the call to implement inquiry-based science instruction at all levels and for all students.

This research was supported by a grant from the National Science Foundation (NSF DUE 0942943, PI: Sacha Kopp), and a grant from the Eunice Kennedy Shriver National Institute of Health and Child Development (5 R24 HD042849) awarded to the Population Research Center at The University of Texas at Austin. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies.

REFERENCES

Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. Cell Biology Education, 4(4), 262–268.

Appleton, K., & Kindt, I. (1999). Why teach primary science? Influences on beginning teachers' practices. International Journal of Science Education, 21(2), 155–168.

Armbruster, P. P. M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. CBE-Life Sciences Education, 8(3), 203–213.

Atwater, M. M., Gardner, C., & Kight, C. R. (1991). Beliefs and attitudes of urban primary teaachers toward physical science and teaching physical science. Journal of Elementary Science Education, 3(1), 3–12.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37(2), 122.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. Proceedings of the National Academy of Sciences, 107(5), 1860–1863.

Bleicher, R. E. (2007). Nurturing confidence in preservice elementary science teachers. Journal of Science Teacher Education, 18(6), 841–860.

Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. Journal of Research on Educational Effectiveness, 1(4), 237–264.

Brownlow, S., Jacobi, T., & Rogers, M. (2000). Science anxiety as a function of gender and experience. Sex Roles, 42(1–2), 119–131.

Bursal, M., & Paznokas, L. (2006). Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science. School Science and Mathematics, 106(4), 173–180.

Cady, J. A., & Rearden, K. (2007). Pre-service teacher's beliefs about knowledge, mathematics, and science. School Science and Mathematics, 107(6), 237–245.

Cobern, W. W., & Loving, C. C. (2002). Investigation of preservice elementary teachers' thinking about science. Journal of Research in Science Teaching, 39(10), 1016–1031.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. American Journal of Sociology 106, 1691–1730.

Coulson, M. R. (1992). Development of an instrument for measuring attitudes of early childhood educators towards science. Research in Science Education, 22(1), 101–105.

Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. Journal of Research in Science Teaching, 42(3), 337–357.

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. Journal of Research in Science Teaching, 51(5), 635–658.

Eagly, A. H., & Chaiken, S. (1993). The psychology of attitudes. Orlando, Florida: Harcourt Brace Jovanovich College.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. Psychology of Women Quarterly, 18, 585–609.

Evans, G., & Durant, J. (1995). The relationship between knowledge and attitudes in the public understanding of science in Britain. Public Understanding of Science, 4(1), 57–74.

Forbes, C. T. (2011). Preservice elementary teachers' adaptation of science curriculum materials for inquiry-based elementary science. Science Education, 95(5), 927–955.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. et al. (2014). Active learning increases student performance in science, engineering, and mathematics. Paper presented at the Proceedings of the National Academy of Sciences, 2013, 19030.

Gess-Newsome, J., & Lederman, N. G. (1999). Examining pedagogical content knowledge. Dordrect, The Netherlands: Kluwer.

Goldberg, F., Robinson, S., Otero, V., Kruse, R., & Thompson, N. (2008). Physical Science and Everyday Thinking, 2nd edition. Armonk, New York: It's About Time, Herff Jones Educational Division.

Haefner, L. A., & Zembal-Saul, C. (2004). Learning by doing? Prospective elementary teachers' developing understandings of scientific inquiry and science teaching and learning. International Journal of Science Education, 26(13), 1653–1674.

Harlen, W. (1997). Primary teachers' understanding in science and its impact in the classroom. Research in Science Education, 27(3), 323–337.

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. Journal of Research in Science Teaching, 49(3), 333–362.

Hopko, D. (2003). Confirmatory factor analysis of the math anxiety rating scale-revised. Educational and Psychological Measurement, 63, 336.

Howes, E. (2002). Learning to teach science for all in the elementary grades: What do preservice teachers bring? Journal of Research in Science Teaching, 39(9), 845–869.

Jarrett, O. S. (1999). Science interest and confidence among preservice elementary teachers. Journal of Elementary Science Education, 11(1), 49–59.

Jussim, L., & Eccles, J. S. (1992). Teacher expectations II: Construction and reflection of student achievement. Journal of Personality and Social Psychology, 63, 947.

Kanter, D. E., & Konstantopoulus, S. (2010). The impact of a project based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry based practices. Science Education, 94(5), 855–887.

Kelly, J. (2000). Rethinking the elementary science methods course: A case for content, pedagogy, and informal science education. International Journal of Science Education, 22(7), 755–777.

Kohut, A., Keeter, S., Doherty, C., & Dimock, M. (2009). Scientific achievements less prominent than a decade ago: Public praises science; scientists fault public, media. Washington, DC: The Pew Research Center for the People and the Press.

Liang, L. L., & Gabel, D. L. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers. International Journal of Science Education, 27(10), 1143–1162.

Ludwig, R., Chimonidou, A., Barr, A., Baumann, M., Dunn, D., English, P., et al. (2013). Hands-on-science: Hands-on, integrated natural sciences for pre-service elementary teachers. Paper presented at the National Association for Research in Science Teaching, Rio Grande, PR.

Mallow, J. V. (1981). Science anxiety: Fear of science and how to overcome it. New York: Van Nostrand Reinhold.

Mallow, J. V., & Greenburg, S. L. (1983). Science anxiety and science learning. Physics Teacher, 21(2), 95–99.

McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in. children's differential response to teacher expectations. Journal of Applied Social Psychology, 32, 159–184.

Mulholland, J., & Wallace, J. (1996). Breaking the cycle: Preparing elementary teachers to teach science. Journal of Elementary Science Education, 8(1), 17–38.

National Research Council. (2000). Inquiry and the national science education standards: A guide for teaching and learning. Washington, DC: National Academies Press.

National Research Council. (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. In S. R. Singer, N. R. Nielsen, & H. A. Schweingruber (Eds.), Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012b). A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. Washi ngton, DC: National Academies Press.

Nelson, G. (2008). Physics and everyday thinking as a model for introductory biology and geology courses. Paper presented at the PTEC-Northwest Regional Conference, Seattle, WA.

Palmer, D. (2002). Factors contributing to attitude exchange amongst preservice elementary teachers. Science Education, 86(1), 122–138.

Park Rogers, M. A., & Abell, S. K. (2008). The design, enactment, and experience of inquiry-based instructon in undergraduate science education: A case study. Science Education, 92(4), 591–607.

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C. et al. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. Journal of Research in Science Teaching, 43(5), 467–484.

Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. College Station, Texas: STATA Press.

Ramey-Gassert, L., Shroyer, M. G., & Staver, J. R. (1996). A qualitative study of factors influencing science teaching self-efficacy of elementary level teachers. Science Teacher Education, 80(3), 283–315.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. American Educational Research Journal, 50(5), 1020–1049.

Seymour, E. (2002). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. Science Education, 86(1), 79–105.

Seymour, E., & Hewitt, N. M. (1997). Talking about leaving: Why undergraduates leave the sciences (Vol. 12). Bo ulder, CO: Westview Press.

Skamp, K. (1991). Primary science and technology: How confident are teachers? Research in Science Education, 21(1), 290–299.

Smith, D. (2000). Content and pedagogical content knowledge for elementary science teacher educators: Knowing our students. Journal of Science Teacher Education, 11(1), 27–46.

Tosun, T. (2000). The beliefs of preservice elementary teachers toward science and science teaching. School Science and Mathematics, 100(7), 374–379.

Udo, M. K., Ramsey, G. P., & Mallow, J. V. (2004). Science anxiety and gender in students taking general education science courses. Journal of Science Education and Technology, 13(4), 435–446.

vanAalderen-Smeets, S. I., Walma van der Molen, J. H., & Asma, L. J. F. (2012). Primary teachers' attitudes toward science: A new theoretical framework. Science Education, 96(1), 158–182.

Volkmann, M. J., Abell, S. K., & Zgagacz, M. (2005). The challenges of teaching physics to preservice elementary teachers: Orientations of the professor, teaching assistant, and students. Science Education, 89(5), 847–869.

Westerback, M. E. (1984). Studies on anxiety about teaching science in preservice elementary teachers. Journal of Research in Science Teaching, 21(9), 937–950.

Westerback, M. E., & Long, M. J. (1990). Science knowledge and the reduction of anxiety about teaching Earth science in exemplary teachers as measured by the science teaching state-trait anxiety inventory. School Science and Mathematics, 90(5), 361–374.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. Contemporary Educational Psychology, 25(1), 68–81.

Xie, Y., & Shauman, K. A. (2003). Women in science: Career processes and outcomes (Vol. 26, No. 73.4). Cambridge, MA: Harvard University Press.

Science
Education

Exploring Teacher Intervention in the Intersection of Digital Resources, Peer Collaboration, and Instructional Design

TORUNN AA. STRØMME, ANNIKEN FURBERG
Department of Teacher Education and School Research, University of Oslo, P.O. Box 1099, Blindern, 0317 Oslo, Norway

Received 15 April 2014; accepted 14 April 2015
DOI 10.1002/sce.21181
Published online 10 June 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This paper reports on a case study of the teacher's role as facilitator in computer-supported collaborative learning (CSCL) settings in science. In naturalistic classroom settings, the teacher most often acts as an important resource and provides various forms of guidance during students' learning activities. Few studies, however, have focused on the role of teacher intervention in CSCL settings. By analyzing the interactions between secondary school students and their teacher during a science project, the current study provides insight into the concerns that teachers might encounter when facilitating students' learning processes in these types of settings. The analyses show that one main concern was creating a balance between providing the requested information and supporting students in utilizing each other's knowledge and understanding. Another concern was balancing support on an individual versus group level, and a third concern was directing the students' attention to coexisting conceptual perspectives. Most importantly, however, the analyses show how teacher intervention constitutes the pivotal "glue" that aids students in linking and using coexisting aspects of support such as peer collaboration, digital tools, and instructional design. *Sci Ed* **99:**837–862, 2015

Correspondence to: Torunn Aa. Strømme; e-mail: t.a.stromme@ils.uio.no

INTRODUCTION

The aim of the current study is to provide insight into teachers' concerns when facilitating students' learning processes in computer-supported collaborative learning (CSCL) settings. Numerous digital learning environments and resources have been developed with the aim of introducing students to scientific concepts (Linn & Eylon, 2011; Quintana et al., 2004). In keeping with this accelerating development, many science classrooms have begun using digital learning resources. Often these digital resources are used in educational settings where students solve open-ended tasks in collaboration with peers and with a teacher who actively guides and participates in the students' development of conceptual understanding.

Several studies have provided valuable knowledge about how to support students' learning processes through use of digital tools (Rutten, van Joolingen, & van der Veen, 2012; Smetana & Bell, 2012), peer collaboration (Howe, Duchak-Tanner, & Tolmie, 2000; Mercer, 2004), and various instructional designs (Linn & Eylon, 2011; Scardemalia & Bereiter, 2006). In most of this research, the analysis focuses on the impact of one or two forms of support. In naturalistic classroom settings, however, various forms of support are present at the same time, which implies that students' learning processes take place at the intersection of different and often coexisting forms of intended support. In addition, in settings where students engage in computer-supported activities, the teacher most often acts as an important resource, providing different forms of guidance during the students' learning activities. Although there seems to be general agreement that teacher support is crucial in computer-supported learning settings, few studies have analytically scrutinized its specific role, especially in CSCL settings (Greiffenhagen, 2012; Urhahne, Schanze, Bell, Mansfield, & Holmes, 2010; Webb et al., 2009).

The current study adds to this body of research by focusing on teacher interventions that support students' development of conceptual understanding in interactions that take place at the intersection of digital resources, peer collaboration, and applied instructional design. To demonstrate the complexity of facilitating students' development of conceptual understanding in these types of settings, we have performed detailed analyses of student and teacher interactions during a student project. In this case study, upper secondary school students designed virtual models of carbon dioxide (CO_2) friendly houses based on scientific theories about energy supply and heat loss from low-energy buildings.

Our analysis focuses on conceptually oriented talk (Furberg, Kluge, & Ludvigsen, 2013), sequences in which the students' and/or teacher's attention is directed to making sense of conceptual issues or, in this case, their talk about heat transfer. Our analytical focus is guided by our interest in exploring the concerns encountered by teachers in settings where students' development of conceptual understanding takes place at the intersection of digital resources, peer collaboration, and instructional design. We analyze student–teacher interactions using van de Sande and Greeno's (2012) conceptualization of "perspectival framing." This perspective enables a combined focus on the participants' social organization during their interaction and how they make sense of conceptual issues.

Research on Support of Students' Conceptual Understanding

Several researchers have pointed out that few studies focus on the role and significance of teacher intervention in CSCL settings (cf. Greiffenhagen, 2012; Urhahne et al., 2010; Webb et al., 2009). Based on analyses of teacher–student interactions in a naturalistic CSCL setting, Greifenhagen (2012) explored teachers' focus in interactions with students during group-work activities. The study reported that teacher interventions targeting conceptually oriented issues, also known as "pedagogical aspects," are intertwined with teacher

interventions targeting classroom management issues. Other studies have focused on the effects of teacher intervention in CSCL settings, and these studies have shown positive effects on students' conceptual understanding when the teacher provides indirect intervention, for instance by prompting questions or encouraging students to retrieve science-based information instead of providing descriptive explanations or prompting fact-based student responses (Hakkarainen, Lipponen, & Järvelä, 2002). Furthermore, a study on students' help-seeking behavior in CSCL settings showed that students sought less help but showed higher learning gains when the teacher provided consolidation instructions in the form of introductions to new tasks, evaluations, and discussions of results in plenary sessions (Mäkitalo-Siegl, Kohnle, & Fischer, 2011).

Our review of studies that have focused on aspects of support other than teacher intervention showed that the studies emphasized one or more of the following aspects: *digital resources, peer collaboration*, and *instructional design*. The majority focused on how various digital resources or tools embedded in computer-based inquiry environments could support student learning. Examples of digital resources are dynamic or static visualizations, computer simulations, interactive tasks, collaboration- and argumentation-supporting tools, domain-specific text, etc., designed to represent a scientific phenomenon and/or central scientific concept (Bell, Urhahne, Schanze, & Ploetzner, 2010; de Jong et al., 2012; Linn & Eylon, 2011). Several studies reported positive effects on students' learning as a result of engaging with various types of computer-mediated representations such as simulations (Rutten et al., 2012; Smetana & Bell, 2012), multiple representations (Ainsworth, 2006), and virtual labs (Baltzis & Koukias, 2009; Kozma, 2003; Zacharia, 2007). In these studies, student learning was primarily measured using pre- and posttests. Despite the consensus on the positive effects of digital support tools on student learning, some studies have also reported challenging findings. For instance, students often have difficulty seeing relationships between different representations of the same phenomenon (van der Meij & de Jong, 2006) or tend to focus on the surface features instead of the underlying scientific principles (Ainsworth, 2006).

Other studies have focused on the influence of peer collaboration in computer-supported settings. Research based on various learning perspectives has emphasized the advantages of peer collaboration in enhancing student learning (Howe et al., 2000; Linn & Eylon, 2011; Mercer, 2004; Scardemalia & Bereiter, 2006; Stahl, 2006). For instance, several studies have found that peer collaboration helps students develop scientific argumentation skills (Linn & Eylon, 2011; Littleton & Howe, 2010), conceptual understanding (Bell et al., 2007; Howe et al., 2007; Linn & Eylon, 2011), inquiry learning skills (van Joolingen, de Jong, & Dimitrakopoulou, 2007), and productive disciplinary engagement (Clark & Sampson, 2007; Engle & Conant, 2002). However, studies have also revealed challenging aspects of peer collaboration. Student talk and collaboration must be cultivated over time, and researchers have pointed to the importance of students learning to deal with disagreements and opposing views on scientific explanations or the problem to be solved (Howe et al., 2000; Mercer, 2004).

Other studies have focused on the impact of the instructional design on student learning processes. A common feature of design-based research is a focus on computer tools or task interventions whose design is informed by idealized models of productive learning. Various instructional models have been developed based on socioconstructivist theories of learning, such as "knowledge building" (Scardemalia & Bereiter, 2006), "progressive inquiry learning" (Muukkonen, Hakkarainen, & Lakkala, 1999), and "knowledge integration" (Linn & Eylon, 2011). Another instructional design model based on similar ideas is the jigsaw model (Aronson, Bridgeman, & Geffner, 1978; Brown et al., 1993), which was the instructional design used in the current study. By breaking classes into groups and

assignments into pieces, the jigsaw model organizes classroom activity to make students dependent on each other to succeed. Several studies have documented positive effects of the jigsaw method on students' learning compared to more traditional teacher-centered and individualized methods (Doymus, Karacop, & Simsek, 2010; Karacop & Doymus, 2013; Tarhan & Sesen, 2012). However, as with all instructional designs, studies have also reported lower or equal academic performance by students under the jigsaw condition compared to more traditional work forms (Hänze & Berger, 2007; Souvignier & Kronenberger, 2007; Zacharia, Xenofontos, & Manoli, 2011).

To summarize, although many studies on science learning in computer-based settings have provided valuable knowledge to the field, we nevertheless stress the value of taking a different analytical approach to provide deeper insight into the role of teacher intervention in these types of settings. In most science classrooms where digital tools and learning environments are used, the teacher orchestrates the support aspects of digital resources, peer collaboration, and instructional design to facilitate students' development of conceptual understanding. By taking an ecological perspective that focuses on teacher interventions taking place at the intersection of digital resources, peer collaboration, and an applied instructional design, and by performing detailed analysis of student–teacher interaction over time, this study aims to provide deeper insight into concerns encountered by the teacher in CSCL settings.

Approaching the Role of Teacher Intervention From a Sociocultural Perspective

Seen from a sociocultural perspective, the teacher holds an important position in students' learning processes (Furberg & Ludvigsen, 2008). First, by virtue of being a scientific expert, the teacher acts as an important conceptual resource for the students. However, the teacher also holds an important position as the facilitator of the learning activities and the instructional design (Squire, MaKinster, Barnett, Luehmann, & Barab, 2003). In addition, the teacher becomes a provider of institutional practices and norms (Mehan, 1991; Mercer, 2004) reflected, for instance, in the assessment criteria, which include expectations regarding how to participate in group work, how to behave in front of a teacher, or how to solve a task appropriately. The relationship between teacher intervention, the tools in use, peers, and instructional design is interdependent: They each influence students' conceptual development in the activity setting. In other words, students' conceptual understanding develops at the intersection of these aspects (Säljö, 2010).

From a sociocultural perspective, learning is seen as a dynamic and dialogical meaning-making process between interlocutors (Linell, 2009; Vygotsky, 1978; Wertsch, 1991). Through their interactions, participants try to interpret and make sense of situations, actions, and scientific concepts. At the same time, the participants make their own interpretations visible and observable to other participants. In this sense, language is seen as the most important tool for making sense of the world, human practices, and ideas and as a tool that mediates thinking and reasoning (Vygotsky, 1986). Talk and discourse are therefore conceived of as a "social mode of thinking" (Mercer, 2004).

Meaning is dialogically constituted in specific practices, and meaning-making involves complex interactions among people, resources, and the organization of the setting (Stahl, 2006). An important part of human conduct and learning processes is the use of various material tools (Säljö, 2010). These can be seen as cultural artifacts that store knowledge and social practices developed over generations (Cole, 1996). This interpretation implies that digital learning environments—often containing representations such as graphs, visualization models, or simulations—are developed to display and represent experts' knowledge

about objects, processes, or phenomena. Students interact with the knowledge and practices stored within digital learning environments when they utilize these representations in their learning activities (Säljö, 2010). In this sense, digital learning environments, such as the SCY-Lab with its embedded digital tools, can be seen as resources for students' development of conceptual understanding.

When engaging with science, students are asked to make sense of diverse concepts. Scientific concepts do not embody fixed or universal meanings but come with historic "meaning potentials" that need to be elaborated on and made relevant to students (Linell, 2009). However, this does not imply that students can come up with just any explanation for a scientific concept. All science domains have cultural contexts that include commonly expressed understandings and ways of talking about conceptions, implying that some ways of representing and talking about scientific concepts are seen as more "correct" or valid than others (Wertsch, 1991). From this perspective, teachers facilitating students' learning processes in computer-supported collaborative settings enforced by various instructional designs must do more than just provide instructional support; they must also orchestrate coexisting support aspects, each with its own affordances and constraints.

The aim of the study is to contribute to the conceptualization of the complexity of teacher intervention within computer-supported learning activities. With an analytical focus on teacher interventions at the intersection of digital resources, peer collaboration, and instructional design, we address the following research question:

RQ: What concerns does the teacher encounter in student–teacher interactions when facilitating students' development of conceptual understanding in CSCL settings?

RESEARCH DESIGN

Design of Learning Activities and Resources

The data in this paper were produced during an intervention study as part of the Science Created by You (SCY) project. The current study is informed by ideas from design-based research (Collins, Joseph, & Bielaczyc, 2004). The objective is to examine interaction and learning in a naturalistic setting but, at the same time, to also study the influence of specific design principles. We used a sociocultural design–based approach; the main difference between this approach and a more "traditional" design-based approach is the status of the design principles in the empirical analysis of the activities and/or learning that takes place during the design experiment (Krange & Ludvigsen, 2009). For instance, in Collins and colleagues' (2004) design-based approach, the design principles are used as the basis both when designing a learning environment and when evaluating the effectiveness of the intervention. In contrast, a sociocultural design–based approach implies that design principles are used in designing learning activities; however, the same design principles are not used as an analytical framework when analyzing the activities and interactions taking place during the intervention. This ensures that the concerns of the participants and their actual activities are scrutinized—not only the researchers' intentions and predefined interests.

Central to the project was the development of the computer environment, the SCY-Lab, which contains various science-related learning modules (de Jong et al., 2012). In the current empirical setting, students were to learn about energy supply and heat loss, and their main task was to design a virtual model of a CO_2 friendly house based information from a variety of resources such as textbooks, Internet-mediated sources, and a heat loss simulation tool embedded in the SCY-Lab. Using the simulation tool, the students calculated the heat loss

TABLE 1
Overview of Project Activities

Day #	Organization	Activity
Day 1	Plenary session	Lecture about energy supply and heat loss from low-energy buildings by visiting expert
	Basic groups	Group task on concept map related to energy supply and heat loss
Day 2	Expert groups	Group 1: Heat loss and insulation
	(Jigsaw model)	Group 2: Heat pumps
	Teacher lecture	Group 3: New renewable energy
	in each field	Group 4: Solar energy
Day 3	Basic groups	Peer-group presentations of individual expert fields
Day 4 + 5 + 6	Basic groups	Design and construction of virtual, CO_2-friendly house with the use of heat loss simulation tool
Day 7 + 8	Basic groups	Preparation for the group presentation
Day 9	Plenary session	Group presentation

of the construction materials used in the virtual house model. The concepts of heat loss (J) and heat transfer coefficient (W/m^2K) were central in the curriculum design. Heat is central to the school science curriculum and is frequently brought up in public discussions about the use of renewable energy in the construction of buildings and private homes.

The participants were 42 upper secondary school students, aged 16–17 years, and two teachers from two general science classes. The two teachers, both in their 10th year of practice, were recruited by the school's principal based on their experience and competence as professional teachers. The project was carried out in 20 school lessons, 45 minutes each, over the course of 2 weeks (see Table 1 for an overview of the project schedule). The design experiment took place at a school situated in Oslo, Norway, as part of the standard instruction schedule.

The SCY-Lab environment was developed by an international project team consisting of programmers, teacher educators, and educational scientists within the SCY project. The design experiment was planned and executed by our local research group. The overall aim of the design experiment was to create a learning setting where we could explore and analyze students' development of conceptual understanding as they use digital learning resources, combined with an instructional design aimed at probing conceptually oriented peer interaction that also included teacher intervention in the form of group guidance. The instructional design and learning activities were planned in collaboration with the two teachers. During this planning phase, the researchers emphasized the significance of peer interaction in the form of conceptually oriented discussions and group-oriented teacher intervention, but the teachers were not given specific instructions on how to facilitate peer interaction and group-oriented teacher intervention. During the design experiment, the teachers, as professional practitioners, had full responsibility for implementing the instructional design without interference from the observing researchers.

Instructional Design, Student Work Forms, and Teacher Intervention

The instructional design was informed by the jigsaw model (Aronson et al., 1978; Brown et al., 1993). This model organizes classroom activity in such a way that students within the same group become experts in different fields. Student collaboration is common in the participating school; however, the particular work form of jigsaw-based instruction used

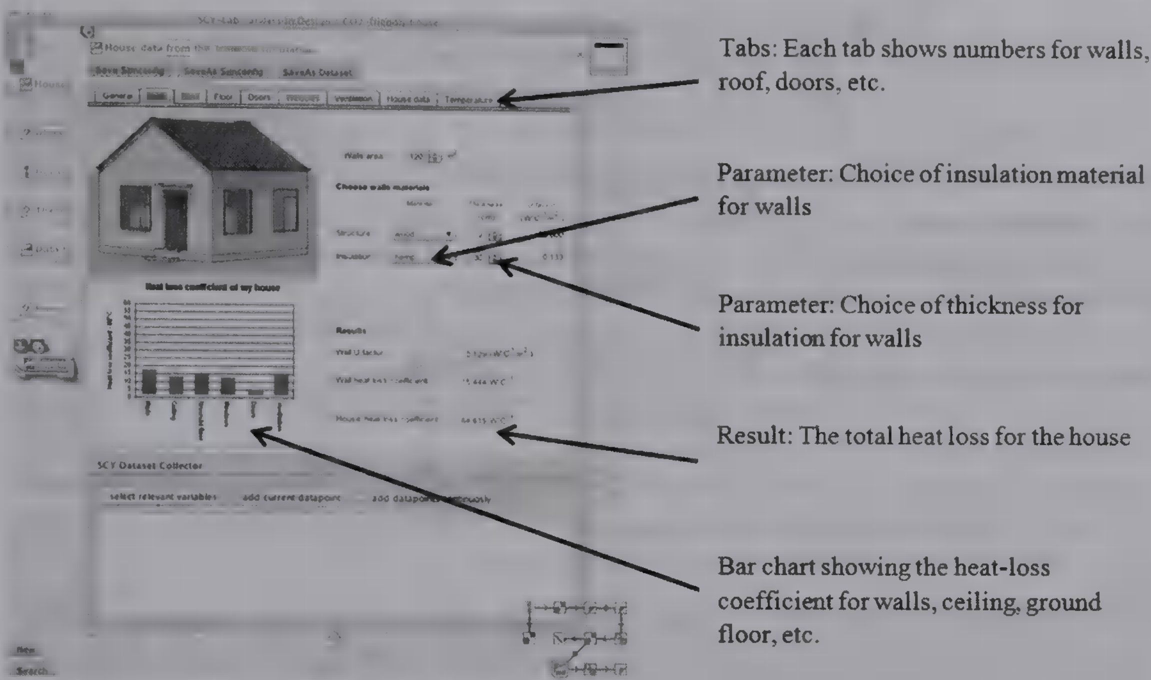


Figure 1. The heat loss simulation tool in SCY-Lab.

in this case was new to the students. Central to the instructional design were the "expert group" sessions during three school lessons at the very beginning of the project. The expert groups, each consisting of three to five students, were given one of four designated "expert fields" to focus on: "heat loss and insulation," "heat pumps," "solar panels and solar thermal collectors," and "new renewable energy." A teacher lectured the expert students in each assigned field. After listening to the teacher, each expert group was asked to produce a one-page written account of the expert topic; the students then reorganized themselves into new groups (termed "basic groups") consisting of one student from each of the four expert groups, and each expert was presented his or her topic of expertise to his or her peers. The goal of the activity was for all students in the groups to gain insight into all expert fields. After the presentations, the groups were asked to design their own virtual, CO_2 friendly house models to present to their class at the end of the project. During the project, the teachers circulated among all the student groups.

The Heat Loss Simulation Tool in the SCY-Lab

A central tool in the SCY-Lab for introducing the students to the concepts of heat transfer coefficient and heat loss was the heat loss simulation tool (see Figure 1), which the students used to calculate how the different construction materials would affect the total heat loss for each house element.

The heat transfer coefficient and heat loss are complex concepts and can be understood from several perspectives. In this study, the teacher explicitly advocated two different perspectives on heat loss. One perspective is the phenomenon perspective (later referred to as "phenomenon framing"): that is, an understanding of heat referring to the thermal energy transferred from one system with a higher temperature to another system with a lower temperature. The second perspective was the formula perspective (later referred to as "formula framing"), in which calculating the heat requires the capacity to see the relation between this concept and other concepts (i.e., power [W] and energy [J])—concepts that, in themselves, can be seen as complex for students. The formula for calculating heat loss

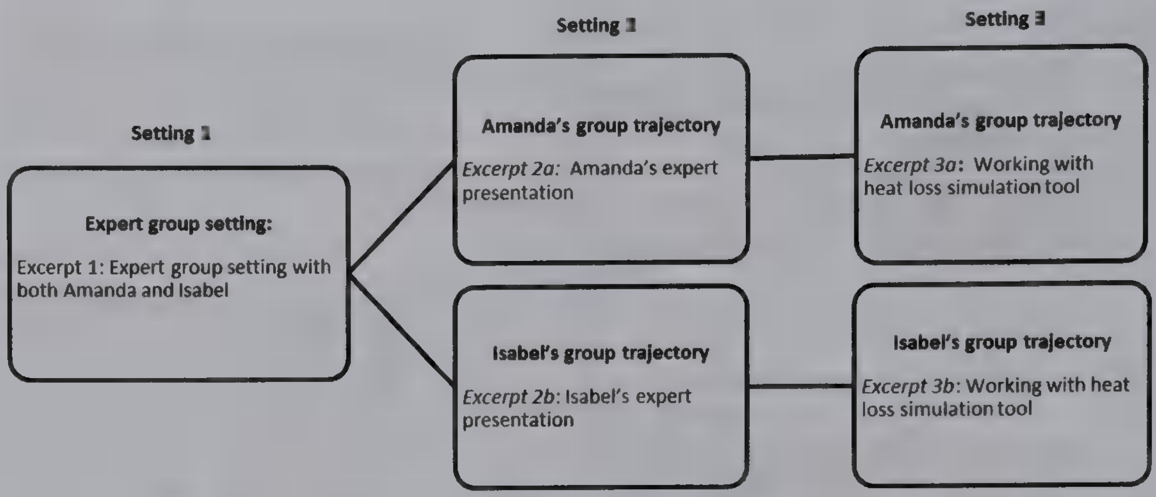


Figure 2. The figure shows the situations from which the excerpts are taken.

is related to the concept of heat transfer coefficient, which is defined as the rate of heat transfer through a building element per square meter per degree of temperature difference (W/m^2K). The engineering notion for the heat transfer coefficient is the U-factor. The concept of U-factor was used in the simulation, and, thus, the students and teachers used the engineering notion when they talked about the heat transfer coefficient.

Data and Analytical Procedure

Three focus groups of four students each were videotaped during the project. The three groups were selected with the teachers' help, based on the criterion of being verbally active. According to the teachers, the students were average- to high-level achievers in science. Our data consisted of 40 hours of transcribed video recordings of the focus groups' interaction, along with field notes taken during classroom observation that were used to contextualize the data.

In this case study, we performed detailed analyses of two students' interactions with their respective peer groups and the teacher. Our analysis focuses on two students, Isabel and Amanda, and how they, together with their peer groups and the teacher, make sense of the concept of heat transfer coefficient. As shown in Figure 2, five interaction excerpts were selected from the two students' interaction trajectories and then analyzed in detail. In accordance with our focus on the role of teacher intervention, we selected excerpts from settings where the teacher engaged with the student groups. Amanda and Isabel participated in the expert group on "heat loss and insulation," and the first analyzed excerpt is from this expert group. In the second part of the analysis, we follow Amanda and Isabel in their two separate basic groups, first in a setting where they present the information and experiences from their expert group session and then in a group-work setting in which the students were to design a virtual house model.

We focused on the interactions between Amanda, Isabel, and their two respective peer groups for several reasons. These two students and their peers were verbally active students. Furthermore, a conceptual topic in Amanda's and Isabel's expert group sessions—the heat transfer coefficient—appeared several times during their basic group discussions as well as in student–teacher interactions. This ongoing verbalized activity in the two groups made the students' development of conceptual understanding transparent in such a way that we are able to analyze in detail how their understanding of heat transfer coefficient developed in the intersection of teacher intervention, digital resources, peer collaboration, and instructional design. Another reason for focusing on these two students and their peer groups is that the

two groups' discussions and work forms differ greatly from one another. Consequently, a dual focus on both Amanda and Isabel and their respective groups enables us to address *variations* in students' development of conceptual understanding, as well as variations in how the teacher intervened.

By analyzing the selected chronological excerpts of the students' interaction trajectory, we are able to show the evolving development of the students' conceptual understanding as well as the opportunities and challenges of teacher intervention in these types of settings. We use the notion of *interaction trajectory* to refer to the analysis of interactions over time (Furberg & Arnseth, 2009; Ludvigsen, Rasmussen, Krange, Moen, & Middleton, 2011). By exploring students' interaction trajectories, we can investigate the changes that take place in students' sense making of the specific domain content as well as how different support aspects influence their sense-making processes. In addition to detailed examinations of specific interaction excerpts, we used ethnographic information documented in video recordings and field notes as a background resource for describing the educational setting. In the discussion and conclusion, we tie our analytic generalizations back to the larger corpus of data, analysis of the extracts, our theoretical grounding, and the literature review.

We used the analytical procedure of interaction analysis, which implies that talk and interaction between interlocutors are analyzed sequentially (Furberg et al., 2013; Jordan & Henderson, 1995). This means that each utterance in a selected sequence is understood and seen in relation to the previous utterance in the ongoing interaction. This practical guideline for analysis supports the idea that analytical descriptions are oriented toward interactional achievements and not what might be taking place in individuals' minds (Linell, 2009).

In our analysis of the student–teacher interactions, we also use a set of analytical concepts on "perspectival framing" adopted from van de Sande and Greeno (2012). Here, framing refers to the way in which participants understand the activity in which they are engaged. We specifically focus on two interrelated aspects of framing: the first aspect, "conceptual framing," refers to the way in which participants, in this case the students and the teacher, organize information by bringing it to the foreground or background of their attention when they try to achieve mutual understanding of a concept or problem. In the current study, by making use of this concept of framing we are able to show which aspects of heat students attend to. For instance, the students may relate to the concept of heat by foregrounding the phenomenon of heat loss, which in this case is how to isolate a house to minimize heat loss and how heat is transferred through different types of materials as a result of a temperature difference between two systems (phenomenon framing). Students may also work with the concept of heat by foregrounding the formula, which in this case is how to calculate heat loss for different building materials using the heat transfer coefficient (formula framing). A central issue of the participants' development of mutual understanding is what van de Sande and Greeno called "alignment of conceptual framing," which refers to whether the participants interactionally develop common ground and "achieve mutual understanding" of how to organize information when solving a task.

The second aspect of framing, "positional framing," concerns the way participants understand themselves and one another in interactions, "especially regarding the contributions each of them is entitled, expected, and perhaps obligated to make in the group's activity" (van de Sande & Greeno, 2012, p. 2). In small-group settings, students collaborate to solve the task by adopting specific positional framings: "source" and "listener." The source is the person or object that provides information another person needs to understand the issue at stake, and the listener tries to interpret the source for mutual understanding.

By using the analytical concepts of perspectival framing, i.e., positional framing and conceptual framing, we are able to show that social processes and individuals' development of conceptual understanding are intertwined. At the same time, the analytical concepts

make it possible to identify how each individual contributes in the mutual development of conceptual understanding. In turn, using these analytical concepts provides deeper insight into the complexity encountered by the teacher in supporting students' development of conceptual understanding.

RESULTS

The excerpts analyzed here are from three subsequent sessions in the project during which the participants discussed heat transfer. In the initial project phase, the instructional design was based on the previously described jigsaw model. The students were organized in expert groups specializing in a particular field and prepared a manuscript to present to their peers in the basic groups. The expert group analyzed below specialized in heat and heat loss, and the expert group session started with the teacher lecturing on heat transfer and insulation of low-energy buildings. In his lecture, the teacher explicitly emphasized the two conceptual framings of heat transfer coefficient and heat loss: phenomenon framing and formula framing. After the lecture, the students focused on group-work activities and browsed the Internet for relevant information to include in the manuscript. During the group-work activity, the teacher circulated among the groups and engaged in their discussions. Below, five interaction excerpts are analyzed. The first setting is from the expert group focusing on heat loss and insulation, in which Amanda and Isabel participated. Subsequently, we follow Amanda's and Isabel's interaction trajectories as they split up and go back to their basic groups to share the information and experiences from the expert group session.

Setting 1: The Expert Group Session: Unpacking the Heat Transfer Coefficient Formula

In the following episode, Isabel and Amanda, and their peers Mia, Magnar, and Lisa, are preparing for their individual basic group presentations. The students are sitting around a table with their laptops in front of them. Mia has summoned the teacher and asked him to read their manuscript. Thus far, the students have written about how to keep heat inside the house. After reading the manuscript, the teacher points out that they need to include the concept of heat transfer coefficient. Picking up on the teacher's suggestion, Amanda asks the teacher to explain, and we enter the discussion when the teacher is about to give his explanation.

Excerpt 1 (see Table 2) begins with the teacher using a simplified example to explain the formula for calculating the power needed to heat a house with fixed dimensions. Amanda and Isabel follow up with specific inquiries. Using the responses provided by the teacher, the three collectively unpack the heat transfer formula by building on each other's input (lines 5, 7, and 9). Amanda's use of the conclusive term "so" in line 11 indicates that she has come to some kind of understanding, and for the first time she tries out a more cohesive verbalized explanation of the heat transfer formula. The teacher confirms Amanda's statement by nodding. Isabel's immediate response in line 13, opening with the discourse marker "but," indicates she finds something is inconsistent or difficult to understand. However, instead of explicating what this is, she withdraws by saying "just kidding." The teacher, Amanda, and the other students do not prompt Isabel to explain her concerns.

If we look at Isabel's contributions in the rest of the excerpt, it becomes clear that at this point she withdraws from providing conceptually oriented queries and inferences. Amanda, however, continues to provide inferences to which the teacher responds and confirms (lines 14 and 16). In line 18, Amanda states that she understands, to which Isabel adds somewhat humorously, that Amanda, who has explicitly expressed her understanding, can take on

TABLE 2
Excerpt 1

1.	Teacher	Let's say you have 400 square meters of wall, ceiling, and floor in the house
2.	Amanda	Yes
3.	Teacher	And the mean value of the U-factor [*heat transfer coefficient*] for the entire house is one. That is, in order to keep a stable temperature inside the house, which is one degree higher than outside the house, you will need a 400-watt electric heater
4.	Isabel	Oh, my God! ((*yawning and leaning backwards*))
5.	Amanda	400 watts (.) What do you mean? 400 watts of what?
6.	Teacher	A 400-watt heating supply inside
7.	Isabel	Because it's 400 square meters? ((*sits upright again*))
8.	Teacher	400 square meters and the mean value for the U-factor is one--
9.	Isabel	And then you'll need one watt per square meter
10.	Teacher	((*nods*))
11.	Amanda	So, it's like watts multiplied by um (.) no, no (.) The size is multiplied by the U-factor in order to find out how much wattage we need?
12.	Teacher	((*nods*))
13.	Isabel	But here ((*points to the computer screen containing her notes from the teacher's lecture*)), you found two different things then. Because here you found--No, I was just kidding
14.	Amanda	So, in order to find that U-factor, you take the watt--
15.	Teacher	But this applies to each degree temperature difference between inside and outside
16.	Amanda	So, if there is a difference of 10 degrees, you'll need 400 times 10 watts? Four thous--
17.	Teacher	4000 watts ((*nods*))
18.	Amanda	But, then I think I understand it
19.	Teacher	That's great
20.	Isabel	Great, Amanda. Then you can write the manuscript ((*laughs*)). No, I am just kidding
21.	Teacher	If you're able to explain this to the others, that would be excellent
22.	Amanda	Because if the U-factor is low, you might not need as many watts as well
23.	Teacher	Right, and then you can use less energy in heating
24.	Amanda	Then you save more electrical energy
25.	Teacher	((*nods*))
26.	Amanda	Oh, yes, then I understand it. We need to write that down ((*pointing at Mia who is writing the manuscript*))
		((*The teacher leaves the room, and the students continue working on their joint manuscript. When the students write about the U-factor in the manuscript, Amanda is the one who dictates what Mia writes*))

Transcript notations: [] Text in square brackets represents clarifying information = Indicates the break and subsequent continuation of a single utterance ? Rising intonation : Indicates prolongation of a sound Underlined: Emphasis in speech (.) Short pause in speech (# of seconds) The time, in seconds, of a pause in speech [. . .] Utterances removed from the original dialog - Single dash in the middle of a word denotes that the speaker interrupts herself -- Double dash at the end of an utterance indicates that the speaker's utterance is incomplete ((Italics)) Annotation of nonverbal activity.

the job of finishing their manuscript. The teacher picks up on Isabel's shift in focus, and emphasizes once more the importance of explaining the heat transfer coefficient to their peers. The episode ends with Amanda checking another specific detail with the teacher, before focusing on what they need to include in their manuscript (line 22).

By applying van de Sande and Greeno's concept of positional framing to Excerpt 1, we can highlight two distinctive aspects of the participants' contributions. The first aspect concerns changes in the participants' positional framing: changes in who is providing information (the source) and who is requesting and interpreting information (the listener). Our analysis of the interaction shows that from the beginning the teacher took the source position by responding to the students' inquiries about the heat transfer coefficient. Amanda and Isabel took the position of constructive listeners by posing inferences and inquiries along the way. Isabel's withdrawal toward the end, however, can be seen as a change in her positioning from a constructive listener to a more passive listener. Amanda undergoes a more substantial shift in positional framing, toward taking the source position. When Amanda provided cohesive reasoning about the heat transfer coefficient and stated she understood, her peers, voiced by Isabel, suggested that she should be responsible for writing the part about the heat transfer coefficient in their document. In other words, Isabel invoked Amanda as a possible source. Amanda's utterance toward the end of the excerpt signals that she acknowledged and took on the appointed role as source when she asserted that they needed to put the things they had talked about in the manuscript.

The second analytical aspect concerns the teacher's elicitation of the students' understanding, or a lack thereof. In the opening, the teacher responded to both Isabel and Amanda's concluding inferences. From the point where Isabel withdrew, though, the teacher's main attention was on Amanda. In addition, the teacher did not pick up on Isabel's query when she signaled that she saw inconsistencies in their joint reasoning. Furthermore, the teacher did not prompt the other students to explicate their understanding. These interactions in Excerpt 1 show the challenges that most teachers face in group-work settings: balancing supporting an individual student's understanding with the group's mutual understanding. As we will see in the following, the variations in the students' understanding of the concepts had consequences for the interactions in both Amanda and Isabel's basic groups in which the two, in the role of expert students, were to provide a detailed explanation of the concept heat. In Excerpts 2a and 2b, we follow Amanda in her basic group.

Amanda's Interaction Trajectory in Her Basic Group

Setting 2a: Amanda's Expert Presentation. In Excerpt 2a (see Table 3), the expert students are back in their basic groups where they are to provide a short introduction to their designated expert topic. Amanda is the last one in her group to present. In terms of conceptual framing, Amanda approaches heat loss and insulation within two conceptual framings: first within phenomenon framing by explaining the importance of insulation for keeping the heat inside the house and then within formula framing, when she explains how to calculate the heat transfer coefficient. During her presentation, the teacher enters the room quietly. Standing in the background, he listens to Amanda's presentation. We enter the setting when Amanda is about to finish her presentation.

The excerpt starts with Amanda giving a complex and somewhat imprecise account of heat and the heat transfer coefficient. For instance, she uses the domain-specific terms watts, joules, heat, and heat transfer coefficient without elaborating on their meaning and provides vague formulations and explanations such as "release the U-factor (heat transfer coefficient)" and "the U-factor (heat transfer coefficient) is the way you calculate power" (line 1). However, regardless of Amanda's dense and unelaborated account of the heat

TABLE 3
Excerpt 2a

1.	Amanda	The U-factor [*heat transfer coefficient*] is the way you calculate how many watts are needed in order to keep the house warm and how much insulation and such. The U-factor is the watts divided by meters squared multiplied by the temperature difference. [. . .] Then you will find the number of kilojoules being released, and then you know that you at least need so many watts in order to keep the heat inside. And preferably more watts than that. And that also affects insulation. If you have bad insulation, then you will release a lot more U-factor, right. And therefore you will need a lot more electrical power. Did you get it? So, you see the connection, don't you?
2.	Linnea	Yes ((*yawning*))
3.	Ole	Yes
4.	Amanda	You understood this, right? It isn't very complicated. You only have to change and switch the formula when you want to find the different numbers and values. [. . .] Yes, this is really all I had ((*smiling*))
5.	Ole	Then we are finished? ((*looking at the teacher*))
6.	Teacher	What have you learned? ((*looking at Ole*))
7.	Ole	Learned and learned. Like (2) like, there are practical solutions, for ventilation and such, that I didn't know about how it functions, and that it was a rather smart thing with the hot air inside that heated outside air coming in. That was quite logical, but I didn't know that
		[. . .]
8.	Teacher	The U-factor, did you understand any of that? It's a difficult concept to understand in a way ((*looking at all the students*))
9.	Linnea	I did at least learn something about it
10.	Teacher	In such a way that you are able to see it as more than a number?
11.	Amanda	I think I was able to explain it quite well
12.	Teacher	That's great ((*giving a thumb's-up sign*))
		((*The teacher leaves the students after a short conversation about what the next task will be*))

transfer coefficient, Linnea and Ole explicitly confirm their understanding when Amanda asks if they have understood what she has explained (lines 2 and 3). Ali, however, remains silent. Sensing that the students consider themselves ready to move on to another task, the teacher interrupts and asks Ole what he has "learned" by listening to Amanda's presentation (line 6). Ole responds to the teacher's question by using the phrase "learned and learned"[1] (in Norwegian, *lært og lært*), which can be interpreted as a way of expressing that he has *perceived* some of the things Amanda explained, which is not the same as *understanding* everything she said (line 7). Then Ole gives an example of something he did understand, which was the part about heat recovery ventilation. After listening to Ole's account of the recovery ventilation, the teacher asks if the students understood what Amanda said about the heat transfer coefficient and adds that this is a complex matter (line 8). Linnea responds, "I did at least learn something about it," indicating that she, like Ole, understood some of the things that Amanda explained to them, but also that some parts were harder to grasp (line 9). Again, the teacher provides an understanding-oriented request and emphasizes the

[1]The phrase *lært og lært* represents what is termed an X-och-X construction in Swedish (Lindström & Linell, 2007), which is also used in Norwegian. This is a reactive pattern: Repeating a previously used term twice signifies that the previous utterance was not quite adequate.

importance of seeing the U-factor as more than just a value. Before any of the addressed students answer, Amanda interjects with a positive validation of her own performance, to which the teacher provides a positive appraisal and leaves (line 12).

By focusing on the participants' positional framing, i.e. their positioning as sources that provide information or as listeners that are requesting and interpreting provided information, we can highlight some concerns encountered by the teacher and students. In this setting, Amanda had the designated position of an expert on heat, a position she accepted. Focusing first on Amanda's peers, the absence of follow-up questions combined with the students' ambiguous utterances about what they have "learned" can be seen as evidence that they found it difficult to relate to their expert peer, as well as an expression of their difficulty with challenging their expert peer to provide a better or more extensive explanation.

Turning the focus to the teacher's positional framing, the analysis shows that in this setting the teacher placed himself quietly in the background when Amanda was presenting. He did not interrupt her presentation, and he did not interfere by providing elaboration or supplementary information, even if he might have perceived that the other students were uncertain. Instead, he limited himself to directing the students' attention toward focusing on the heat transfer coefficient, along with asking them whether they had understood the concept. In other words, when the teacher refrained from taking a source position, he was left in the middle, neither a source nor a listener. This situation seems to be a double-edged sword in that the teacher risked undermining the expert student's role as the designated source if he took the source position. However, by allowing Amanda's peers to "get away with" stating their (partial) understanding instead of making them accountable for displaying their understanding, the teacher put himself in a position in which he was incapable of knowing what the students did and did not understand.

Concerning the participants' conceptual framing, Excerpt 2a shows that the teacher attempted to emphasize the importance of both phenomenon framing and formula framing. Both framings were addressed by Amanda in her presentation. When prompted to account for what they had learned from Amanda, her peers mainly provided phenomenon framings of heat. Consequently, the teacher's method of directly prompting Amanda's peers about their understanding of the heat transfer coefficient was a way of confirming that the students' attention was directed not only at phenomenon framing but also at formula framing.

Setting 3a: Working With the Heat Loss Simulation Tool. Before the following excerpt (see Table 4), Amanda and Ali had worked for a while with the simulation tool in the SCY-Lab. This tool explicitly addresses the heat transfer coefficient and helps students calculate it for different building elements for their virtual house. Ali and Amanda browse the Internet for information about the Norwegian requirements for house insulation. When the teacher enters the room, Ali seizes the opportunity to ask the teacher about the variation in different materials' heat transfer coefficients. While the two talk, Amanda continues browsing the Internet for information.

The excerpt begins with Ali wanting to know whether a *high* or *low* heat transfer coefficient value indicates the best heat loss result, since he observed from interacting with the simulation tool that steel has a much higher heat transfer coefficient than wood (lines 1 and 3). The teacher responds, "Steel conducts heat very well." Ali seems to interpret the teacher's statement "conducts heat very well" as a positive quality and infers that steel would be a better choice than wood for the exterior material (line 5). This implies that Ali infers that a high heat transfer coefficient is validated as better than a low coefficient, and, consequently, materials with a high heat transfer coefficient are better to use for insulation.

TABLE 4
Excerpt 3a

1.	Ali	Steel has a U-factor [*heat transfer coefficient*] of 1000
2.	Teacher	Uhum
3.	Ali	And wood and such have only four. Why is there such a big difference?
4.	Teacher	Steel conducts heat very well
5.	Ali	So, it's better with steel then? ((*referring to steel being a better exterior material in their house*))
6.	Teacher	No, you know, if you've got steel going through from the inside to the outside, then a thermal bridge, as one calls it, will appear because steel conducts heat very well. That is possible to feel for yourself if you've got a matchstick. It can burn all the way until the flame reaches your finger without getting very warm. If you take a nail, metal, you know.
7.	Ali	Uhum
8.	Teacher	And warm it at the end, then it won't take long before it has conducted the heat so much that you're not able to hold it
9.	Ali	Yea, that's true. But, what does high and low U-factor mean? Is it good with a high or low U-factor?
10.	Teacher	Amanda, what is best: a high or low U-factor?
11.	Amanda	Low is better. ((*Keeps looking at the computer screen*))
12.	Teacher	Uhum
13.	Ali	Then, it's better with wood than steel?
14.	Teacher	Uhum. The U-factor is a measure of energy flow ((*The conversation changes to another topic*))

The teacher picks up on Ali's incorrect inference and responds by explaining about thermal bridges: How a piece of metal gets warm very quickly when exposed to a flame. Ali's response signals that he understands the teacher's explanation of how steel is a better heat conductor than wood, but that he still grapples with determining whether a high or low heat transfer coefficient is considered the best when it comes to insulation quality (line 9). Instead of answering Ali's straightforward question, the teacher bounces the question over to Amanda, the designated expert on heat and insulation. Amanda responds that "low" is better, to which Ali infers that wood must be better than steel. The teacher confirms Ali's inference, and adds that the heat transfer coefficient measures energy flow.

In terms of the participants' positional framing in this setting, we see that in the opening of the excerpt, the teacher took the source position when he responded to Ali's inquiries about the meaning of high and low heat transfer coefficients. Ali took the listener position. Toward the end of the excerpt, however, the teacher redirected Ali's question to Amanda (line 9). By doing this, the teacher withdrew from the source position, just as he did in the last basic group setting (Excerpt 2a), and at the same time he invoked Amanda, the expert student, as the source.

Another aspect of the positional framing in this excerpt concerns the simulation tool's position as a source. The simulation was designed to help students understand the relevance of calculating the heat loss of insulation materials and to help them unpack the role of the heat transfer coefficient in the formula for calculating heat loss. Thus, the simulation supports a formula framing of heat. Ali's focus on the values calculated according to the formula shows that in this setting he foregrounded the formula framing. Furthermore, the interaction in Excerpt 3a (see Table 4) shows that the simulation did not provide enough support for Ali to understand the heat transfer coefficient or interpret high and low values

for this coefficient. When responding to Ali's queries, the teacher explained by pointing to what happens to steel when it is exposed to flame. In other words, by using the steel example to explain the differences in materials' heat transfer coefficients, the teacher used a phenomenon framing of heat to explain heat from a formula framing. Based on the teacher's linking of conceptual framings, Ali then correctly concluded that wood is better than steel for exterior use.

Before we end the analysis of Amanda and her peer's group work, we will describe their conceptual framings of heat in their plenary presentation at the end of the project. In the presentation, the students emphasized the heat transfer coefficient. Amanda explained how to calculate the heat transfer coefficient, presented values for it, and based her final argument on why their house was a low-energy building on this concept. She compared the house's total heat loss with the requirements for heat transfer coefficients for low-energy buildings. Our interpretation is that formula framing was in the foreground in the students' presentation, whereas phenomenon framing was in the background.

Isabel's Interaction Trajectory in Her Basic Group

Setting 2b: Isabel's Expert Presentation. We enter Isabel and her group's interaction trajectory when Isabel is about to finish her 10-minute expert presentation. She ends by asking if any of her peers have questions. One student asks Isabel to elaborate on the concept of heat transfer coefficient, and in the following excerpt, Isabel is about to reply to this request.

In the opening of Excerpt 2b (see Table 5), Isabel, the designated expert, explains that a low heat transfer coefficient means that the house does not emit much air, and then adds that insulation prevents wind from entering the house (lines 1 and 2). Mary, who is trying to understand, follows up by asking an inferential question. By using the discourse marker "but," she signals she does not understand what Isabel is saying. Mary confirms that she understands what Isabel says about the insulation stopping the wind, but points out that she still wants to know whether the insulation warms up incoming cold air as well as letting the warm air pass into the house (lines 3, 5, and 7). By responding with an initial "No," Isabel signals that Mary's inference is wrong and continues by emphasizing that not all but some of the air will enter the house (line 8). Seemingly unsatisfied with Isabel's answer, Mary repeats her question about whether the insulation warms up the incoming cold air. The tone in her voice indicates that she is getting frustrated. At this point, Elise interjects, and says that she does not understand what they are talking about. The tone of her voice signals that she also is becoming frustrated (line 10). In lines 12 and 14, Isabel tries again to explain how insulation works. In her explanation, she still focuses on how insulation stops wind from entering the house, but she also provides a more elaborated account of ventilation. Mary's question about whether the insulation warms the incoming air remains unanswered. Elise's "si, si" (pronounced with an Italian accent) (line 13) can be interpreted as a signal that she accepts Isabel's explanation without necessarily understanding or agreeing with it. Isabel's wind-stopper explanation remains unchallenged, as Malin (line 15) and Mary (line 17) confirm when Isabel asks if they now understand.

In terms of the participants' positional framing in this setting, we see that Isabel, as the expert on the designated topic, took the source position in this setting and her peers initially took the listener position. Mary's continuous search for an answer and the agitated atmosphere show the difficult position the students were in when the expert student Isabel was unable to provide the requested information. However, when Mary challenged Isabel's idea by presenting an alternative idea, Mary took on a potential source position. This left the group with two potential sources: one arguing for the assumption that the pores in

TABLE 5
Excerpt 2b

1. Isabel	Low U-factor [*heat transfer coefficient*] is like, uhm::: that you don't emit so much air
	[. . . .]
2. Isabel	The air is not supposed to go through, because then the air comes from the outside and in, right? When heavy wind hits the house, it is supposed to, uhm::: the material will stop it. Because that is the reason why it's got many small air uhm::: air holes, right?
3. Mary	Yes, but isn't that like--
4. Isabel	So that it stops.
5. Mary	Yes, it stops, but isn't it like the air in a way meets that insulation, so that the insulation heats up the air that comes in?
6. Isabel	But the air--
7. Mary	And then it's releasing heat to the house, and then it's releasing the cold to like the outward layer of the house. Isn't it like that?
8. Isabel	No, like, if it's heavy wind, all of the air isn't entering the house. But some of it will enter the house.
9. Mary	It will hit the insulation, but the insulation makes it warm instead of cold?
10. Elise	What are we really talking about now?
11. Mary	I don't know. I don't get it. Like, how it happens, like how=
12. Isabel	Well, first the external wall stops most of the air, right, but then there are small- Like there are these tiny loopholes that perhaps only a tenth of it, or something, manages to pass through. And then there is the plastic, right, and then that, what's it called, the insulation material that stops everything. Right?
13. Elise	Si si [*said with Italian accent*]
14. Isabel	And then inside the house you have the ventilation system that circulates the air inside the house. There will always be some draft, right. But mostly around the windows. And the air that passes through, or if you have a window slightly open, the ventilation system will circulate it around the house, right. And then it moves out, and new air enters. Right?
15. Malin	Yes
16. Isabel	Anything else? ((*giggles*))
17. Mary	No. I got it now
18. Isabel	Okay. Good
	((*The students start working on the next task*))

insulation prevent some of the wind from penetrating the insulation and the other arguing for the assumption that insulation transforms cold air into warm air when the air enters the insulation.

These ways of explaining insulation have been documented in several studies that have focused on students' common intuitive ideas (Chu, Treagust, Yeo, & Zadnik, 2012; Clark, 2006; Schnittca & Bell, 2011). In the current study, both versions in the end remained unchallenged. Instead, the group ended up confirming that they accepted Isabel's version. However, their confirmation does not necessarily mean that the students agreed or understood. Their consent might well have expressed that the students wanted to bring the unsettled issue to an end and that they acknowledged the designated expert as the source. Either way, the students ended up settling for a version inconsistent with the scientific conceptions held by experts in the field.

The second analytical point concerns the participants' conceptual framing (i.e., the way in which the participants organize information by bringing it in the foreground or background of their attention). When asked to elaborate on the heat transfer coefficient, a question that is positioned within formula framing, Isabel responded by providing a phenomenon description of wind hitting the insulation (phenomenon framing). Isabel could have responded to the question without repositioning the conceptual framing, for instance by elaborating on how to make calculations using the heat transfer coefficient. However, she chose not to invoke a formula framing. The reason for her choice might be found in the previous analysis of the participants' interaction in the expert group setting (Excerpt 1). This analysis showed that Isabel grappled with understanding how to make calculations using the heat transfer coefficient, and instead focused on something that she found easier to understand and explain to her peers.

Setting 3b: Working With the Heat Loss Simulation Tool. The group has just started calculating the heat loss of their house using the simulation tool. The students have changed several parameters to see the consequences for their house. When the teacher enters the room, the students have still not commented on any changes in output factors in the simulation. Malin seizes the opportunity to ask the teacher how to operate the simulation tool.

In the opening of Excerpt 3b (see Table 6), Malin asks the teacher what they are supposed to do with the simulation (line 1). The teacher takes the mouse cursor and explains in detail how to operate the simulation. Without explaining the term, he tells the students that they are to calculate the heat transfer coefficient of the construction material in their house (lines 2 and 4). At this point, Malin seizes the opportunity to ask the teacher what the heat transfer coefficient is (line 5). Instead of answering the question, the teacher bounces the question over to Isabel, the designated expert on this topic. Isabel replies by providing a definition of heat transfer coefficient (line 7). Malin follows up by asking what value is considered high for a heat transfer coefficient (line 8). Not picking up on Malin's request about the value, Isabel responds by going into the consequences of a high heat transfer coefficient (line 9). Not getting the answer she was looking for, Malin reframes her question. The tone in her voice along with bursting out the imperative "Numbers" shows that she is getting frustrated (line 10).

The conversation continues with a few more similar turns (lines 11–14). Isabel does not provide the information Malin is looking for until Malin asks Isabel a yes-and-no-question, and she confirms that 1 is a high value (line 14). Isabel adds that 0.3 or 0.13 is considered to be very good. The last part of Isabel's reply is formulated as a question addressed to the teacher, and her use of the past tense ("wasn't it?") indicates that she is referring to something they have talked about before, probably in the expert group setting. Instead of confirming Isabel's answer, the teacher encourages the students to look up the values on a Web page made available to them in the SCY-Lab (line 18). However, the students seem to have received the information they needed since they do not look up the links but continue working with the simulation.

In the context of van de Sande and Greeno's perspectival framing, three analytical points can be highlighted. In terms of the participants' positional framing, Malin's way of directing her question directly to the teacher shows that she invoked him as a possible source in this setting. The teacher, however, refrained from taking the source position and handed the question to Isabel, the designated expert student, by invoking her as a possible source. Isabel, accepting the appointed source position, tried to come up with a reasonable answer to Malin's question. The challenge appeared when Isabel did not

TABLE 6
Excerpt 3b

1. Malin	What are we supposed to do?
2. Teacher	Here you can find out how much energy the house uses. And then you choose for each (.) building element. Here are the walls ((*points with the mouse cursor at the relevant tab*)). And then you can choose-- What should the walls be made of?
3. Malin	U::hum
4. Teacher	Structure, that means what they are made of-- So you've got walls of wood, walls of concrete-- [. . .] And, then you have the total U-factor [heat transfer coefficient] for the walls here. ((*Points with the mouse cursor to the calculated value for the U-factor for the walls in the simulation*))
5. Malin	What is the U-factor?
6. Teacher	The U-factor? Isabel learned quite a bit about that. What is the U-factor? ((*looking at Isabel*))
7. Isabel	U::hm That's the unit of measurement for how much heat loss there is in the house per square meter
8. Malin	What is a high U-factor then? ((*looks at Isabel*))
9. Isabel	That is not good. Because then the house emits--
10. Malin	Yes, but what is it? How high is it then?
11. Isabel	Then the house emits much heat--
12. Malin	Number!
13. Isabel	Then it gets cold more easily, and you need to heat it all the time.
14. Malin	But, is like 1 a lot?
15. Isabel	Yes
16. Malin	That is a lot.
17. Isabel	What was it again? 0.3 was really good. That was a super window, wasn't it? No, 0.13 ((*looks at the teacher*))
18. Teacher	If you are to-- If you find one of those links, then they are written there.
19. Elise	Isn't it good that it is- We are not supposed to lose so much heat, or lose so much this? ((*changes a parameter so that the bar showing heat loss in the diagram increases and points at the increasing bar*))
20. Teacher	No, the U-factor should be low
	((*The students carry on their work with the simulation*))

provide the information Malin was looking for and Malin became frustrated as a result. This mismatch between the information requested by Malin and the information provided by Isabel can better be understood as a lack of alignment in the students' conceptual framings (i.e., to what extent do participants achieve a mutual understanding of how to organize information when solving the task). Malin wanted to know the specific value for a high heat transfer coefficient, implying that she foregrounded the formula framing. Isabel, however, answered the question descriptively and focused on why it is desirable to have a high heat transfer coefficient, and in so doing foregrounded the phenomenon framing. Put differently, the students' divergence was caused by an observed but unaddressed lack of alignment in their conceptual framing. This challenge was settled when in the end Isabel provided the information Malin requested, implying that Isabel aligned her conceptual framing with Malin's. However, the two possible ways of framing the concept of heat remained unaddressed and implicit.

The final analytical point concerns the teacher's positional framing. By refraining from taking the source position, as he also did when interacting with Amanda and her basic group peers, the teacher found himself in the middle, positioned as neither a source nor a listener.

As seen before, the teacher faced the challenge of balancing the risk of undermining the expert student's role as the designated source against providing students with information that would help them to continue with their task on their own. The teacher's solution in the current situation was to recommend that Isabel and her peers look for the information on the Internet.

Regarding Isabel's and her peer group's conceptual framings during their presentations at the end of the project, although Isabel's group argued for their choices of materials based on heat loss, they did not explicitly use the concept of heat transfer coefficient during their presentation. This omission may indicate either that the students in this group did not consider the concept particularly relevant for communicating their choices or that they were unsure how to account for the meaning of the concept and therefore avoided mentioning it. Nevertheless, this implies that the phenomenon framing was maintained in the foreground of Isabel's and her peers' presentation, whereas the formula framing was the background, or more or less left out entirely.

DISCUSSION

The overall aim of this study was to provide deeper insight into the complexity of supporting students' development of conceptual understanding in collaborative learning settings. In the following sections, we first discuss the central empirical findings from the analyses of the student–teacher interactions and then discuss the empirical findings in relation to previous research findings.

Our analytical approach used van de Sande and Greeno's (2012) conceptualization of perspectival framing to investigate the participants' interactions. This method directed our analytical attention on what is referred to as the participants' positional framing (i.e., how participants relate to each other in interaction, as source and listener) as well as their ongoing work with constructing and making sense of coexisting conceptual framings (i.e., in what ways the students organize information or how they approach a concept from different perspectives). This analytical approach revealed four major concerns the teacher encountered and had to deal with as he facilitated the students' learning processes.

One concern encountered by the teacher was *directing the students' attention to coexisting conceptual perspectives*. The analyses show how the teacher continuously tried to ensure that the students considered the two conceptual framings, phenomenon framing and formula framing. This balancing activity was observed in the expert group session (Excerpt 1) and the two basic group settings (Excerpts 2a and 3b). Moreover, the analysis showed that the students tended to foreground the phenomenon framing and were more likely to background, or even exclude, the formula framing. We hypothesize that the main reason for the teacher's continuous effort to balance the two framings was that he perceived that the students struggled to explain heat loss within the formula framing, and thus saw that he had to provide additional support in the form of directing the students' attention and discussion toward this more complex issue.

The second concern encountered by the teacher was *creating a balance between providing the requested information versus supporting students in utilizing each other's knowledge and understanding*. The interaction analyses show that the teacher used two positioning strategies. One strategy was to take the source position, implying that he provided the information the students requested and needed. The teacher used this strategy in the expert group session (Excerpt 1). The second positioning strategy was refraining from taking the source position combined with designating other potential sources. This strategy was mainly used in the basic group settings in which the teacher tended to respond to the students' queries by invoking the designated expert students and bouncing the questions

over to them (Excerpts 3a and 3b). In cases where he discovered that the expert student was incapable of providing the requested information, he invoked other potential sources, such as the designated Web resources. This implies that the teacher adjusted his positional framing strategy depending on the setting. He willingly took the source position within the expert group setting, whereas he refrained from the same position in the basic group settings.

So, how can the teacher's choice of the two positional framing strategies be explained? We argue that the teacher's choices of strategies must be seen in relation to the jigsaw design. This instructional design required that all students be given roles as experts and novices, i.e., intended source and listener positions. Being in the source position was challenging for the students, but being in the listener position was also challenging, since the students found it hard to challenge or ask for elaborations of the explanations provided by the expert students, i.e., the designated source (Excerpts 3a, 2b, and 3b). The teacher's decision to refrain from taking the source position in the basic group settings can be seen as a way of supporting the student in the expert position, and also a way of sustaining the main intention of the instructional design: to facilitate shared understanding by means of conceptual input from all students in their roles as experts and novices. This demonstrates the challenge that the teacher faced in balancing his positional framing. By taking the source position in the basic group setting, he risked undermining the students' exercise of their designated roles as experts. However, by refraining from taking the source position, he put himself in a position where he became incapable of knowing what the students did and did not understand.

The third concern encountered by the teacher was *balancing individual and group support*. This challenge was especially prominent in the expert group setting (Excerpt 1). Here, the teacher's attention was mainly directed toward responding to one student's queries and inferences, causing the other students to withdraw from being constructive listeners, i.e., refraining from engaging in an effort to achieve mutual understanding. This implies that the teacher seemed to provide sufficient and productive support to one individual student, but at the same time missed out on the opportunity to provide support that benefitted all the students in the group. Furthermore, the analyses show that the teacher tended to prompt the students to state, but not to display, their understanding (Excerpts 1 and 2a). This implies that the teacher did not know whether the group or the individual group members had achieved a sufficient understanding of the concept in focus or whether the students held intuitive ideas, as was the case in one of the student groups.

The fourth concern encountered by the teacher was *enabling the students see the relevance of the simulation*. In the empirical setting, the students engaged with a simulation tool designed to support their understanding of making calculations using the heat transfer coefficient (Excerpts 3a and 3b). The analyses of the participants' interaction while they engaged with the tool demonstrate the considerable interpretative effort needed for the students to make sense on different levels: making sense of what to do with the simulation (Excerpt 3b), understanding the term heat transfer coefficient as well as its relative value (Excerpts 3a and 3b), and seeing the relevance of the heat transfer coefficient in a broader sense. The simulation tool apparently did not provide enough conceptual support for the students to achieve a mutual understanding of the concept of heat transfer coefficient, since both groups summoned the teacher. In this sense, the teacher became an important resource in this setting by explaining instructions and directing students toward supplementary sources, as well as working with the students' conceptual ideas of the heat transfer coefficient.

The empirical findings of the current study confirm, as well as supplement, findings from previous research that have focused on student learning in computer-supported collaborative

settings. This study provides deeper insight into peer interaction in these types of settings. Previous studies have documented productive aspects of peer collaboration: for example, it can foster learning-promoting talk and interaction among students (cf. Howe et al., 2007; Stahl, 2006). However, studies have also shown the challenging aspects of peer collaboration—for instance, that students rarely engage in discussions characterized by "constructive listening" (van de Sande & Greeno, 2012) or "exploratory talk" (Mercer, 2004), and that collaboration as an activity is difficult for students (Furberg & Arnseth, 2009). Our analyses show how these types of settings open up possibilities for peer-driven elicitation of intuitive ideas and attempts to develop mutual conceptual understanding. However, the analyses also demonstrate the significance of an intervention by a teacher who explicates coexisting intuitive ideas and scientific ideas, as well as settling potential conceptual disagreements. In addition, the analyses show the challenging aspects of peer collaboration and the importance of regulative support provided by the teacher to ensure that all students get a chance to provide their understanding. This support also includes elicitation of students' intuitive ideas or misconceptions.

Turning the focus to the instructional design, several studies have scrutinized the productive effects on students' construction and sharing of scientific arguments within settings based on various jigsaw designs (Aronson et al., 1978; Brown et al., 1993; Karacop & Doymus, 2013). Nevertheless, studies have also reported on more modest, or even negative, effects of jigsaw designs (Hänze & Berger, 2007; Souvignier & Kronenberger, 2007; Zacharia et al., 2011). Regarding the jigsaw design, the current study yielded differing findings. On the productive side, the jigsaw design supported an environment that urged ongoing, conceptually oriented peer discussions and elicitation of ideas. A challenging aspect of the jigsaw design, however, is the variations in the students' conceptual framing (what aspects of heat the students chose to focus on), the students' conceptual understanding, and the quality and accuracy of the explanations provided by the expert students. The peer interaction analyses show that these variations had a huge impact on the conceptually oriented discussions in the basic groups. Furthermore, the analyses show that even if an expert student had developed an understanding of a concept or was capable of providing a sophisticated explanation of the concept at issue, this did not ensure that mutual understanding developed in the peer group. Another challenging aspect of the jigsaw design is related to the participants' designated positions as sources and listeners. The instructional design with its designated positions was challenging for the teacher as well; for instance, the teacher sometimes needed to refrain from taking on the source position when the students obviously grappled with understanding the concepts at issue to avoid undermining the students' designated roles as experts.

Several studies have reported positive effects on students' learning related to the use of digital simulations (Rutten et al., 2012; Smetana & Bell, 2012). The analyses in the present study demonstrated productive sides of the students' use of the heat loss simulation tool; for instance, it prompted conceptually oriented talk in both groups related to unpacking the heat transfer coefficient. The simulation also became a tool for the teacher in that it actualized and supported his emphasis on the importance of understanding heat not only within phenomenon framing but also within formula framing. However, the analyses also show the considerable interpretive effort needed for students to make sense of the concepts embedded in the digital representations, in this case the concept of heat transfer coefficient. This finding coincides with previous process-oriented studies that focused on how digital representations are invoked and made sense of by students in collaborative learning settings (Furberg et al., 2013). Most importantly, the study shows the significance of teacher intervention in elaborating, explaining, and contextualizing the concepts and scientific principles embedded within digital resources. This interpretive support is important for students' development

of conceptual understanding; it also shows the potential of digital support resources in prompting conceptually oriented teacher–student talk.

We began by drawing attention to a characteristic feature of the undertaken research within the field of CSCL: The analytical focus favors the impact of one or sometimes two support aspects. The core argument forming the basis for the current study was the importance of applying an ecological perspective: seeing students' learning processes as intertwined with support provided by a teacher, peer collaboration, instructional design, and their engagement with digital resources in use. To explore the "intertwinedness" and the concerns encountered by the teacher in these types of settings, we have argued for, and demonstrated, the relevance of opening up these classroom practices by means of interaction analysis (Furberg et al., 2013; Jordan & Henderson, 1995). Furthermore, we have argued for the significance of an analytical attention on *interaction trajectories*, which implies following interactions over time, across settings and between students groups (Ludvigsen et al., 2011). The main finding of the current study is that teacher interventions are crucial in supporting students' development of conceptual understanding. The teacher's intervention constitutes the "glue" in the setting by providing support in the intersection of peer collaboration, digital resources, and instructional design; when something goes awry in the intersection of these various forms of support, the teacher becomes the last layer of support.

Concluding Remarks

Facilitating students' development of conceptual understanding in CSCL settings is not a trivial task for teachers. Students' capabilities to participate critically, but constructively, in peer discussions, elicit and explore each other's intuitive ideas and scientific thinking, and settle disagreements are skills that needed to be cultivated over time. Research has shown the value of training students to participate in scientific discourse combined with introducing discussion ground rules (Howe et al., 2000; Mercer, 2004). To support this development, however, teachers must prioritize spending time and effort cultivating a classroom climate that supports critical, but constructive, exchanges of views, knowledge, and shared conceptual sense making. The current study demonstrates the necessity for teachers to critically scrutinize the productive as well as challenging aspects of any instructional design in relation to how the instructional design supports, or even prevents, productive learning processes. Overall, the findings from the current study show, more than anything, the complexity involved when designing computer-supported learning settings. We argue that teachers will benefit both from being aware of this complexity and from seeing themselves as facilitators of students' learning processes as they take place in the intersection of peer collaboration, digital support tools, and teacher intervention.

As a concluding remark, we will point out that several authors have emphasized the need for studies that address the role of the teacher and also the instructional setting in relation to students' use of computer simulations (Rasmussen & Ludvigsen, 2010; Smetana & Bell, 2012). Although the current study provides a contribution, further studies are needed in these areas. Specifically, studies that focus on different science domains, digital resources, and instructional designs are needed to understand the complexity of students' conceptual sense making in naturalistic CSCL settings.

We would like to thank our colleagues in the MEDIATE research group of the Faculty of Educational Sciences, University of Oslo; participants in the National Graduate School in Education (NATED); and participants in the TELS research group directed by Professor Marcia Linn for constructive feedback on earlier drafts. Especially, the authors would like to thank Professor Sten Ludvigsen, Associate Professor Sonja M. Mork, Associate Professor Ingeborg Krange, and Dr. Alfredo Gil

Jornet for valuable comments on earlier drafts. We would also like to thank the anonymous reviewers who have provided helpful comments on the refinement of the article. This work was funded by the Department of Teacher Education and School Research and InterMedia, University of Oslo. The study was conducted in the context of the Science Created by You (SCY) project, which was funded by the European Community under the Information and Communication Technologies (ICT) theme of the Seventh Framework Program for R&D (grant agreement 212814). This article does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16, 183–198.

Aronson, E., Bridgeman, D. L., & Geffner, R. (1978). Interdependent interactions and prosocial behavior. Journal of Research and Development in Education, 12, 16–26.

Baltzis, K. B., & Koukias, K. D. (2009). Using laboratory experiments and circuit simulation IT tools in an undergraduate course in analog electronics. Journal of Science Education and Technology, 18(6), 546–555.

Bell, T., Shcanze, S., Graber, W., Slotta, J. D., Jorde, D., Berg, H. B., & ... Evans, R. H. (2007). Technology-enhanced collaborative inquiry learning: Four approaches under common aspects. In R. Pintó & D. Couso (Eds.), Contributions from science education research (pp. 451–463). Dordrecht, The Netherlands: Springer.

Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. International Journal of Science Education, 32(3), 349–377.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Cmpione, J. (1993). Distributed expertise in the classroom. In G. Sloman (Ed.), Distributed cognitions (pp. 188–288). New York, NY: Cambridge University Press.

Chu, H.-E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of students' understanding of thermal concepts in everyday contexts. International Journal of Science Education, 34(10), 1509–1534.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. Cognition and Instruction, 24(4), 467–563.

Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. International Journal of Science Education, 29(3), 253–277.

Cole, M. (1996). Cultural psychology: A once and future discipline. Cambridge, MA: Belknap Press.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. The Journal of the Learning Sciences, 13(1), 15–42.

de Jong, T., Weinberger, A., Girault, I., Kluge, A., Lazonder, A. W., Pedaste, M., & ... Ney, M. (2012). Using scenarios to design complex technology-enhanced learning environments. Educational Technology Research & Development, 60(5), 883–901.

Doymus, K., Karacop, A., & Simsek, U. (2010). Effects of jigsaw and animation techniques on students' understanding of concepts and subjects in electrochemistry. Educational Technology Research and Development, 58(6), 671–691.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. Cognition and Instruction, 20(4), 399–483.

Furberg, A., & Arnseth, H. C. (2009). Reconsidering conceptual change from a socio-cultural perspective: Analyzing students' meaning making in genetics in collaborative learning activities. Cultural Studies of Science Education, 4, 157–191.

Furberg, A., Kluge, A., & Ludvigsen, S. (2013). Student sensemaking with science diagrams in a computer-based setting. Computer Supported Learning, 8, 41–64. doi:10.1007/s11412-013-9165-4

Furberg, A., & Ludvigsen, S. R. (2008). Students' meaning-making of socio-scientific issues in computer mediated settings: Exploring learning through interaction trajectories. International Journal of Science Education, 30(13), 1775–1799.

Greiffenhagen, C. (2012). Making rounds: The routine work of the teacher during collaborative learning with computers. International Journal of Computer-Supported Collaborative Learning, 7(1), 11–42.

Hakkarainen, K., Lipponen, L., & Järvelä, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall, & N. Miyake (Eds.), CSCL 2: Carrying forward the conversation (pp. 129–156). Mahwah, NJ: Erlbaum.

Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. Learning and Instruction, 17(1), 29–41.

Howe, C., Duchak-Tanner, V., & Tolmie, A. (2000). Co-ordinating support for conceptual and procedural learning in science. In R. Joiner, K. Littelton, D. Faulkner, & D. Miell (Eds.), Rethinking collaborative learning (pp. 81–100). London, England: Free Association Books.

Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., & ... Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. Learning and Instruction, 17(5), 549–563.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. The Journal of the Learning Sciences, 4(1), 39–103.

Karacop, A., & Doymus, K. (2013). Effects of jigsaw cooperative learning and animation techniques on students' understanding of chemical bonding and their conceptions of the particulate nature of matter. Journal of Science Education and Technology, 22(3), 86–203.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. Learning and Instruction, 13(2), 205–226.

Krange, I., & Ludvigsen, S. R. (2009). The historical and situated nature design experiments: Implications for data analysis. Journal of Computer Assisted Learning, 25(3), 268–279.

Lindström, J., & Linell, P. (2007). X-och-x som samtalspraktik och grammatisk konstruktion. [X-and-x as a conversational practice and grammatical construction]. In E. Engdahl & A. M. Londen (Eds.), Interaktion och kontext [Interaction and context] (pp. 19–89). Lund, Sweden: Studentlitteratur.

Linell, P. (2009). Rethinking language, mind and world dialogically: Interactional and contextual theories of human sense-making. Charlotte, NC: Information Age.

Linn, M. C., & Eylon, B.-S. (2011). Science learning and instruction: Taking advantage of technology to promote knowledge integration. New York, NY: Routledge.

Littleton, K., & Howe, C. (2010). Educational dialogues: Understanding and promoting productive interaction. New York, NY: Routledge.

Ludvigsen, S. R., Rasmussen, I., Krange, I., Moen, A., & Middleton, D. (2011). Intersecting trajectories of participation: Temporality and learning. In S. R. Ludvigsen, A. Lund, I. Rasmussen, & R. Säljö (Eds.), Learning across sites: New tools, infrastructures and practices (pp. 105–121). New York, NY: Routledge.

Mäkitalo-Siegl, K., Kohnle, C., & Fischer, F. (2011). Computer-supported collaborative inquiry learning and classroom scripts: Effects on help-seeking processes and learning outcomes. Learning and Instruction, 21(2), 257–266.

Mehan, H. (1991). The school's work of sorting students. In D. Boden & D. H. Zimmerman (Eds.), Talk and social structure: Studies in ethnomethodology and conversation analysis (pp. 71–90). Cambridge, England: Polity.

Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. Journal of Applied Linguistics, 1(2), 137–168.

Muukkonen, H., Hakkarainen, K., & Lakkala, M. (1999). Collaborative technology for facilitating progressive inquiry: Future learning environment tools. In C. Hoadley & J. Roschelle (Eds.), Proceedings for: Computer support for collaborative learning. Designing new media for a new millennium: Collaborative technology for learning (pp. 406–415). Mahwah, NJ: Erlbaum.

Quintana, C., Reiser, B. J., Davies, E. A., Krajcik, J., Fretz, E., Duncan, R. G., & ... Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. The Journal of the Learning Sciences, 13, 337–386.

Rasmussen, I., & Ludvigsen, S. (2010). Learning with computer tools and environments: A sociocultural perspective. In K. Littleton, C. Wood, & J. K. Staarman (Eds.), International handbook of psychology in education (pp. 399–433). Bingley, England: Emerald.

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. Computers & Education, 58, 136–153.

Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: Technologies, social memory and the performative nature of learning. Journal of Computer Assisted Learning, 26, 53–64.

Scardemalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), Cambridge handbook of the learning sciences (pp. 97–118). New York, NY: Cambridge University Press.

Schnittca, C., & Bell, R. (2011). Engineering design and conceptual change in science: Addressing thermal energy and heat transfer in eighth grade. International Journal of Science Education, 33(13), 1861–1887.

Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. International Journal of Science Education, 34(9), 1337–1370.

Souvignier, E., & Kronenberger, J. (2007). Cooperative learning in third graders' jigsaw groups for mathematics and science with and without questioning training. British Journal of Educational Psychology, 77(4), 755–771.

Squire, K. D., MaKinster, J. G., Barnett, M., Luehmann, A. L., & Barab, S. L. (2003). Designed curriculum and local culture: Acknowledging the primacy of classroom culture. Science Education, 87(4), 468–489.

Stahl, G. (2006). Group cognition: Computer support for building collaborative knowledge. Cambridge, MA: MIT Press.

Tarhan, L., & Sesen, B. A. (2012). Jigsaw cooperative learning: Acid-base theories. Chemistry Education Research and Practice, 13(3), 307–313.

Urhahne, D., Schanze, S., Bell, T., Mansfield, A., & Holmes, J. (2010). Role of the teacher in computer-supported collaborative inquiry learning. International Journal of Science Education, 32(2), 221–243.

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. Learning and Instruction, 16, 199–212.

van de Sande, C., & Greeno, J. G. (2012). Achieving alignment of perspectival framings in problem-solving discourse. Journal of the Learning Sciences, 21(1), 1–44.

van Joolingen, W. R., de Jong, T., & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. Journal of Computer Assisted Learning, 23(2), 111–119.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). Thought and language. Cambridge, MA: Harvard University Press.

Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). "Explain to your partner": Teachers' instructional practices and students' dialogue in small groups. Cambridge Journal of Education, 39(1), 49–70.

Wertsch, J. V. (1991). Voices of the mind: A sociocultural approach to mediated action. Cambridge, MA: Harvard University Press.

Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits. Journal of Computer Assisted Learning, 23(2), 120–132.

Zacharia, Z. C., Xenofontos, N. A., & Manoli, C. C. (2011). The effect of two different cooperative approaches on students' learning and practices within the context of a WebQuest science investigation. Educational Technology Research and Development, 59(3), 399–424.

Science
Education

Supporting Preservice Science Teachers' Ability to Attend and Respond to Student Thinking by Design

HOSUN KANG,[1] CHARLES W. ANDERSON[2]
[1] *University of California Irvine, 3200 Education Building, Irvine, CA 92697-5500, USA;* [2] *Michigan State University, 319A Erickson Hall, 620 Farm Lane, East Lansing, MI 48824, USA*

Received 20 August 2014; revised 22 April 2015; accepted 11 May 2015
DOI 10.1002/sce.21182
Published online 14 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: A teacher's ability to attend and respond to student thinking is a key instructional capacity for promoting complex and deeper learning in science classrooms. This qualitative multiple case study examines 14 preservice science teachers' (PSTs) responses to learning opportunities created to develop this capacity, as provided by a teacher preparation program. The PSTs engaged in multiple cycles of designing assessments and analyzing student work in coordination with clinical experiences in the field. Drawing upon the notions of responsiveness and noticing, we analyze teaching episodes for whether and how the PSTs in this study attended and responded to student thinking in instructional contexts. Several teaching episodes provide evidence of PSTs' productive responsiveness—suggesting modification in specific elements of instructional design to create better conditions for advancing students' scientific thinking. In general, however, the episodes suggest uneven success in PSTs' responses to student thinking. The findings point to two considerations in designing learning opportunities to enhance PSTs' responsiveness: (a) the use of high-quality assessment tasks that make student thinking visible and (b) helping PSTs to reframe the problems by deprivatizing PSTs' interpretations of student responses. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**863–895, 2015

Correspondence to: Hosun Kang; e-mail: hosunk@uci.edu

INTRODUCTION

The science education community aims to prepare future science teachers who are capable of promoting all students' deeper learning. Current scholarship in science education advocates complex and deeper learning, characterized by students' robust reasoning, participation in meaningful scientific practices, and sense-making conversations (NGSS Lead States, 2013; NRC, 2007, 2012). To support this kind of learning, teachers must be able to recognize and build on students' ideas and experiences, use students' ways of reasoning as valuable resources, and continuously adapt instruction in response to both the process and progress of student learning (Bransford, Brown, & Cocking, 1999; NRC, 2012; Sawyer, 2006). Attending and responding to student thinking is one core instructional capacity that makes that form of learning possible in classrooms.

Despite the general consensus that developing this instructional capacity is a worthy goal of preservice teacher education (Interstate New Teacher Assessment and Support Consortium Science Standards Drafting Committee, 2002; Kloser, 2014; NRC, 2010), the teacher preparation community has been struggling to figure out how to best support preservice science teachers (PSTs) in cultivating this capacity. The fundamental challenge resides in the fact that we do not yet have well-developed ideas about how and under which conditions PSTs develop this ability, and how their learning progresses throughout their careers. Currently there is little empirical evidence to inform program design and professional learning opportunities that support PSTs in developing this important instructional capacity during the teacher preparation period (NRC, 2010; Windschitl, 2005).

This study intends to fill some gaps in the literature about preservice teacher learning coordinated with a designed learning opportunity. The PSTs of this program engaged in multiple cycles of designing assessment tasks as a part of planning, implementing plans in their field placement, and collecting, analyzing, and reflecting on student work during their two years in the program. The assessment activities (designing assessment tasks, analyzing student work, and reflecting on their practices) were purposefully designed to draw preservice secondary science teachers' (PSTs) attention to students' scientific thinking and to guide them to respond to it. We examine *how* secondary PSTs responded to this learning opportunity. Specifically, the following research questions (RQs) are addressed:

1. What was the nature of assessment tasks (items) that were designed or selected by the PSTs?
2. How did PSTs interpret student responses? What did PSTs attend to and how did they make sense of it?
3. What form of instructional change did PSTs suggest (or not)?
4. How were PSTs' attention and responsiveness mediated by assessment activities shaped by their interactions with people in contexts?

We begin by discussing how we conceptualize teachers' responsiveness. The review of previous study provides theories of action behind the pedagogical approach (i.e., engaging PSTs in assessment activities) in relation to responsive teaching. Following the details of our research activities, we present the different responses to this learning activity by the participating PSTs. The variations across the PSTs' responsiveness are accounted for and discussed in relation to the affordances and constraints of the pedagogical approach. The article concludes with some implications and recommendations for designing professional learning programs that cultivate the ability to attend and respond to student thinking.

CONCEPTUAL FRAMEWORK

Teachers' Responsiveness and Noticing

Teachers' responsiveness is a complex construct to define because responsive teaching involves attending to and addressing multiple problems emerging within a classroom that is populated with diverse students who have different learning needs. Researchers in the fields of science and mathematics education have conceptualized "teacher responsiveness" differently (Elby et al., 2014), and there exists disagreements about what counts as responsiveness in the literature (see Gay, 2000; Hammer, Goldberg, & Fargason, 2012; Rosebery & Puttick, 1998; Sherin, Jacobs, & Philipp, 2011). In this study, we define teachers' responsiveness as the practices of deliberate and ongoing attention and actions that move student learning forward.

This formulation of responsive teaching rests on two premises. First, students have rich nascent resources for reasoning about and making sense of the world around them, and therefore students, even young children, are capable of engaging in complex reasoning and scientific sense-making when appropriate support is provided (Maskiewicz & Winters, 2012; Metz, 1995, 2004; NRC, 2007, 2012). Second, students experience and learn science meaningfully when they are positioned as competent science leaners and their ideas and experiences are recognized, brought forth, and built upon in a supportive learning community (Bransford et al., 1999; Calabrese Barton et al., 2013; Nasir, Rosebery, Warren, & Lee, 2006; NRC, 2005, 2012). In that sense, teachers' responsiveness to students' ideas and ways of thinking, their lived experiences, and their identities as members of multiple cultural and discourse communities is essential to support all student learning (Gay, 2000; Hammer et al., 2012; Thompson et al. 2016). In this study, we focus on one dimension of a teacher's responsiveness—attending and responding to student thinking.

Research on teachers' attention and responsiveness to student thinking describes responsive teaching as involving three aspects (see Figure 1). In the first step of responsive teaching, teachers purposefully elicit student ideas (see Hammer et al., 2012). From an instructional standpoint, this step refers to teachers' deliberate efforts to create opportunities for students to show what they know with use of certain forms of assessment (Kang, Thompson, & Windschitl, 2014). Next, teachers interpret and make sense of student responses—how and why students respond in this particular way. Researchers note that a teacher's ability to recognize and interpret the connections between students' ideas and the discipline is critical to being responsive to student thinking (Coffey, Hammer, Levin, & Grant, 2012; Levin, Hammer, & Coffey, 2009). In the final step of responsive teaching, teachers take action based upon these interpretations.

Teachers make their pedagogical decisions based on what they attend to within students' ideas and how they interpret the students' understanding. For instance, in looking over student work, a teacher observes that several of her students are confused about nuclei from the lesson where students made observations of two cells under a microscope (i.e., multiple students understand the nucleus to be a cell in and of itself within the larger cell). The teacher interprets the students' confusion as being related to the limited observations of different kinds of cells within the prior lesson. In this way, this teacher problematizes her own initial instructional design (i.e., not providing sufficient opportunities for students to find patterns across the different types of cells); therefore, she modifies the subsequent activities to provide a complementary experience in the following lesson—taking actions to address the perceived problem with regard to student understanding. A responsive teacher continuously engages in this cycle of eliciting, attending, interpreting, and responding to student thinking on the course of instruction at the scale of in-the-moment interaction as well as on a daily basis.

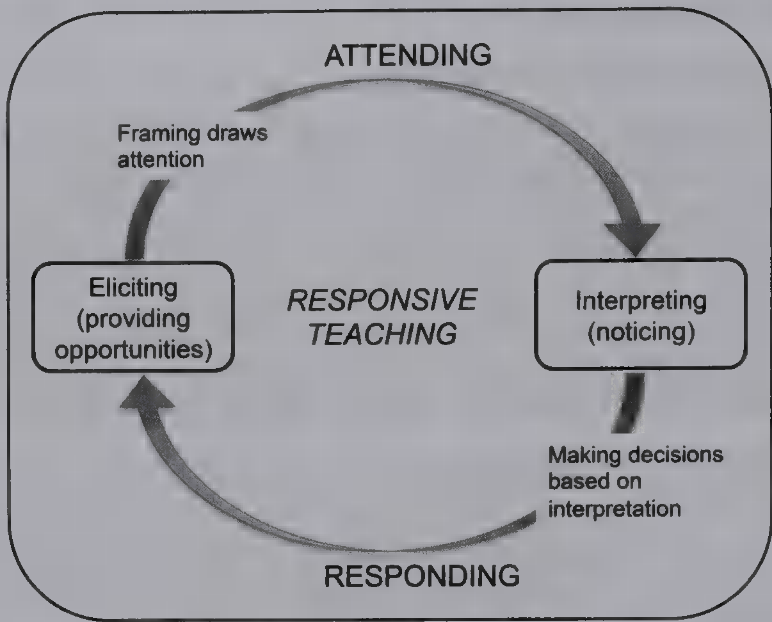


Figure 1. A framework for teacher responsiveness.

Researchers who study the practices of attending and responding to student thinking point out *teachers' ability to notice* as a key mediator that shapes teachers' responsiveness. The assumption is that teachers' pedagogical decisions about actions they take depend on what teachers notice while interpreting students' responses or situations. In mathematics teacher education literature, van Es and Sherin (2005; 2008) identified three key components of teachers' noticing practices. The first component is identifying "what is important" in a teaching situation. When PSTs interpret students' responses to assessments, for example, they "call out" or "highlight" certain information. The highlighted information shows a PST's act of deciding what is noteworthy and deserves further attention. Second, noticing involves using knowledge of the subject matter, of students as science learners, as well as of their local context to reason about events as they unfold. The final aspect of noticing is making connections between specific events and the broader principles of teaching and learning. It requires teachers to categorize and extrapolate from the specific to the general as they respond to the question of, "What is this a case of?" During this process, PSTs (or teachers) connect what they observe in their classrooms to broader principles of teaching and learning, which affects their courses of action.

What one notices is inevitably affected by the structure of expectations about the situation—in other words, the ways in which the situation is framed (Russ & Luna, 2013). A teacher who frames science teaching as working on and with students' ideas is likely to attend to and notice various forms of students' ideas and ways of reasoning, and then act upon them to revise students' thinking. In contrast, a teacher who frames science teaching as delivering canonical scientific knowledge is likely to attend to the correctness of students' responses.

In the fields of sociology (Goffman, 1974), sociolinguistics (Tannen, 1993), and cognitive science (Minsky, 1985; Schank, 1990), framing is conceived as a kind of *schema*, or the "structure of expectations" (Ross, 1975) grounded in one's experience of the world in a given culture (or combination of cultures) (Tannen, 1993). In the field of preservice teacher education, the powerful role of teachers' initial "frame of reference" (Kennedy, 1999) as shaped by their "apprenticeship of observation" (Lortie, 1975) has been well recognized. Kennedy (1999) argues that one important role of preservice teacher education is to change

PSTs' initial frames of reference, thereby allowing the PSTs to see situations differently and thus generate different ideas about how they might respond to these situations.

Taken together, the previous studies suggest that a teacher's responsiveness to student thinking is shaped by what s/he notices during interpretation, and what s/he notices is driven by the structure of the teacher's expectations—the ways in which the situation is framed.

Cultivating Preservice Science Teachers' Capacity for Attending and Responding to Student Thinking via Assessment

PSTs in this program engaged in multiple cycles of structured formative assessment tasks during their field experiences. Researchers point out that formative assessment at its core consists of attending and responding to student ideas and reasoning, with roots in disciplinary activity and goals (Coffey et al., 2012; Levin et al., 2009; Sadler, 1998). In the literature, formative assessment is typically referred to as the process by which teachers use evidence of students' learning to modify their teaching to make it more effective (Black, Harrison, Lee, Marshall, & Wiliam, 2004). Effective formative assessment is distinguished by two key features. One is teachers' genuine attention to and engagement with ideas, continuous with the disciplinary practices science teachers should be working to cultivate (Coffey et al., 2012). The other is teachers' instructional responsiveness as manifested by modification or adaptation of teaching (Black & Wiliam, 1998; Wiliam & Thompson, 2007). Furtak (2012) reminds us that "formative assessment hinges on a criterion of *use*, and when information is not used to improve performance, it is not formative" (p. 1186). Given the nature of the work, engaging PSTs in a formative assessment process can, in theory, provide scaffolded opportunities for them to attend and respond to students' scientific thinking.

In the context of preservice teacher education, teacher educators generally perceive assessment activities as a promising approach to support PSTs' systemic learning for the following reasons. First, students' written responses produced from preplanned assessment tasks can make a variety of student ideas visible (Furtak & Ruiz-Primo, 2008); and, therefore, provide easy access to student thinking. In theory, PSTs can learn about their students as "sense-makers," paying attention to students' ideas and reasoning, and make the needed modification of their practices based on the evidence of students' understanding. Second, collecting and analyzing student work outside the classroom may provide opportunities for PSTs to develop new insights into situations and student learning. Third, PSTs can take time to analyze student responses to plan their actions based on the information that they gain (Atkin, Coffey, Morthy, Sato, & Thibeault, 2005).

Studies that empirically examine PSTs' engagement in assessment activities, however, reveal the depth of challenges in helping PSTs to attend and respond to students' scientific thinking via assessment activities. In a study of 61 PSTs' formative assessment practices in the context of a semester-long practicum-based assignment, Otero and Nathan (2008) found that the PSTs tended to attend to either everyday experience-based ideas (what they call "experience-based conceptions") or to science-based ideas taught in school (i.e., "academic conceptions"). However, PSTs responded only to science-based ideas even when students' everyday experience-based ideas were elicited. Otero (2006) also found that a "get it or don't" conception was commonly used by PSTs when they engaged in formative assessment, with serious impacts on their instructional practices. Much of novice teachers' knowledge about assessment is "underdeveloped" (Maclellan, 2004), which makes it difficult for them to make sense of student responses that require complex reasoning with evidence (Lyon, 2013).

Teaching is a performance, and it is even more difficult to help PSTs develop the ability to respond to student thinking *in action*. Teachers make pedagogical decisions based on their

interpretations of classroom situations (Kennedy, 1999). A teacher's choice of action at any moment is inherently responsive to the perceived situations and recognition of problems that need to be addressed. An important question to teacher educators is whether and how teachers' choice of actions enhances students' learning by creating better conditions for intellectual and social interactions. Productively responding to students' scientific thinking involves what Schön (1983) calls, "reflection-on-action"—teachers spending time exploring why students responded as they did in relation to disciplinary learning goals and the circumstances of provided learning opportunities. It requires PSTs to *reframe* a problem and integrate knowledge about teaching and learning, which was rarely observed when PSTs engaged in assessment activities (Lyon, 2013). Furthermore, the complex nature of student teaching, a context typically populated with multiple perspectives and expectations from both the school and the program, is not always conducive to PSTs learning how to respond productively to students' scientific thinking. In fact, it is often impossible for PSTs to make actual changes in instructional design on the following day or try out new strategies in someone else's classroom.

RESEARCH DESIGN

Study Context

The context for this study was a reform-oriented five-year undergraduate teacher preparation program. The PSTs in this program took four field-based, disciplinary-specific methods courses during the last two years of the program, the senior (fourth) and internship (fifth) years. The historical relationship between the program and local school communities provided relatively strong leeway for the program to structure PSTs' experiences in the field. The program engaged PSTs in mandatory teaching responsibilities and assessment activities at designated times. This institutional context made it possible for the methods course instructors to design deliberate learning opportunities through assignments coordinated with experiences in the field.

PSTs engaged in about eight teaching cycles—planning, enactment, assessment, and reflection—throughout the two years with coaching from instructors and field supervisors. To develop the ability to respond to students' scientific thinking, the teaching cycles included an assessment component with three subactivities: (a) designing assessment tasks, (b) interpreting student responses, and (d) suggesting changes in instruction (see Figure 2). Each phase was guided with tools (e.g., template, rubric) and various scaffolds to draw PSTs' attention to students' scientific thinking and help them to learn how to respond to it. The following section discusses the connection between each activity and aspects of responsiveness illustrated in Figure 1 as well as the underlying assumptions behind this pedagogical approach.

Designing Assessment Tasks—Eliciting. Assessment tasks provide information about what students know and are able to do (Kang et al., 2014). The requirement to design an assessment task forces PSTs to think about the outcomes they would expect to see if they accomplish their instructional goals. The design of the assessment tasks generated from this process inevitably reflects how PSTs frame knowledge and learning of science when they design or select the assessment.

The PSTs of this program either designed or selected two to three assessment tasks at the planning stage of each teaching cycle. Typically, PSTs interacted with methods course instructors, peers, and their mentor teachers during this time to search for resources and ideas for their assessment tasks. The program provided three kinds of support to assist

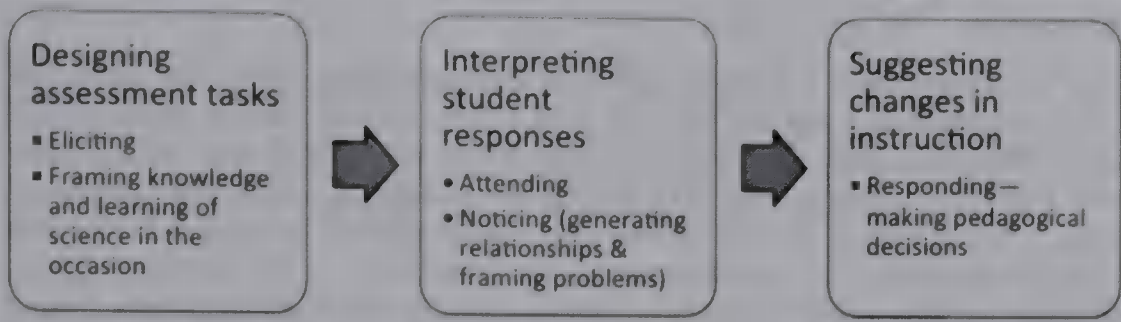


Figure 2. Assessment activities for developing teachers' responsiveness.

PSTs' planning of high-quality assessment tasks that make students' scientific thinking visible. First, the program provided a template that included rubrics about the quality of assessment tasks and detailed prompts to assess students' scientific thinking. In addition, the instructors posted examples of quality assessment tasks on the course website. Lastly, PSTs received individual feedback from the methods course instructors regarding their assessment tasks (items), including suggestions for modifications before enactment.

Interpreting Student Responses—Attending and Noticing. After implementing their lesson or unit plan in the field, PSTs collected and analyzed student work. During this phase, PSTs interpreted student responses while accounting for how and why students produced particular responses. PSTs were expected to attend to and notice various forms of student ideas and thinking such as partial understandings or alternative ideas during this process.

The course instructors required PSTs to select and examine responses from three focus students at different academic achievement levels, except within the last two teaching cycles. The intention was to engage PSTs in an in-depth inquiry about student thinking rather than in superficial levels of analysis, such as "get it or don't get it." In addition, PSTs discussed their analyses of students' responses with their peers and experienced teachers (field supervisors) before they produced a written report. A group of five to six PSTs who were under similar student teaching conditions (subject taught, grade level, school, etc.) collectively looked at the produced student work for about 90 minutes with the facilitation of a course instructor or a field supervisor. Finally, the detailed prompts and rubrics in the template guided PSTs' systematic analysis of students' scientific thinking.

Suggesting Changes in Instruction—Responding. The PSTs suggested changes to their instruction based on their interpretations of student responses while producing a written report. The prompts provided in the report template explicitly instructed PSTs to suggest specific changes to different components of their instructional design that might address the framed problem. The caveat to this approach is that PSTs' suggested changes to instruction did not necessarily reflect the actual instructional adaptation or PSTs' ability to enact them. Further, PSTs might have simply tried to "please" the methods instructors to meet their expectations when interpreting student responses and suggesting changes in instruction. However, it was unlikely that PSTs could respond to students' scientific thinking productively in action without first being able to thoroughly interpret student responses and make deliberate decisions based upon those interpretations.

Participants

This investigation uses a multiple case study approach (Stake, 2004; Yin, 1989). We selected participants in a way that maximized the variation among cases to enhance transferability (Merriam, 2009). In this program, each cohort consisted of about 30 secondary science PSTs. The majority of the PSTs were college students with no formal teaching experience. Typically, about one-third of PSTs had some experience working with scientists in a laboratory setting, either as undergraduate research assistants or in a graduate program. Each cohort included one or two PSTs who chose teaching as their second careers. During the two academic years (2008–10) of this study, we interacted with three cohorts of PSTs who were at either the intern (fifth year) or senior (fourth year) stages. Between four and six PSTs were selected from the volunteers in each cohort based on three criteria: (a) a spectrum of personal backgrounds (e.g., major, gender, content area, teaching experiences, research experiences in science, and career choice); (b) school contexts (e.g., suburban academic-oriented schools vs. urban high-need schools); (c) mentor teachers' relationship with the program. PSTs' personal backgrounds, and in particular their prior teaching experiences, were considered because of their potential influence on the ways in which PSTs frame the work of science teaching and learning. School contexts and mentor teachers' relationships with the program were also considered for the same reason.

In Year 1 (2008–09), a total of eight PSTs were selected from two cohorts (see Table 1). The first group of four PSTs from cohort I were interns, and the second group of four PSTs, from the cohort II, were seniors during Year 1. Two of them continued to participate in the Year 2 (2009–10) as they proceeded to their internship year. One opted out of this study, expressing his difficulties in managing time during his internship year. One was placed at an affluent suburban school for her internship year that was located a long distance from the research site, with new mentor teachers who had not worked with the program. We decided to replace these PSTs with two other interns, Adam and Alice. Adam's mentor was a graduate of the program, and Alice's mentor had been working with the program for several years. In the second year of this study, in addition to Adam and Alice from Cohort II, another group of four PSTs were selected from a new cohort (Cohort III). These PSTs only participated in this study during their senior year.

In addition to PSTs, 12 mentor teachers and two course instructors participated in this study. Among the 12 mentor teachers (three PSTs worked with the same mentor in different years), five graduated from the program and the other four had previously been working with the program as mentor teachers. Three mentor teachers were newly recruited teachers who did not have any relationship with the program. At the end of the year, one of the new mentor teachers, Ms. S., was not recommended to return. Her field supervisor cited management/behavior-oriented teaching and inflexibility as the reasons for the recommendation.

Sources of Data

The primary source of data was 32 *sets of teaching episodes* generated from 32 teaching cycles taught by 14 participating PSTs. A teaching episode included lesson or unit plans, the descriptions of assessment tasks, worksheets or slides that showed the actual assessment tasks (items), samples of student responses, and PSTs' written analysis and ideas for improvement. Each year we asked the participating PSTs to give us at least one teaching video recorded during their teaching episodes. Half of the teaching episodes included teaching videos (one to two teaching videos per one PST). These data provided information

TABLE 1
Profiles of the 14 Preservice Science Teachers

Cohorts	Pseudonym	Year 1	Year 2	Gender	Major	Teaching Experiences	Research Experiences	Career Choice	School Context
Cohort I	Monica	Intern	–	Female	Biology	No	No	–	Suburban HS
	David	Intern	–	Male	Biology	No	No	–	Suburban HS
	Teresa	Intern	–	Female	Biology	An outdoor program	No	–	Urban HS
	Sarah	Intern	–	Female	Chemistry	No	No	–	Urban HS
Cohort II	Leslie	Senior	Intern	Female	Biology	No	No	–	Urban HS at both years
	Shannon	Senior	Intern	Female	Chemistry	No	Yes	Second career	Urban HS Suburban HS
	Mary	Senior	–	Female	Chemistry	No	No	–	Suburban HS
	Kevin	Senior	–	Male	Physics	No	No	–	Suburban HS
	Adam	–	Intern	Male	Biology/physical science	No	Yes	–	Suburban HS
	Alisa	–	Intern	Female	Earth science	No	No	–	Urban MS
Cohort III	Lori	–	Senior	Female	Biology	No	No	–	Urban HS
	Lynn	–	Senior	Female	Biology	No	No	–	Suburban HS
	Stella	–	Senior	Female	Biology	No	Yes	–	Suburban MS
	Scott	–	Senior	Male	Physics	No	No	–	Urban MS

about PSTs' assessment design, interpretations about student responses, and their ideas for instructional modification.

A second major source of data was *interview transcripts*. A semistructured interview was conducted with the participating PSTs individually at the end of each year. During the interview, the interviewer showed a preselected 3–5 minute long segment of the PST's teaching video. Selected clips illustrated students (either individually or in groups) engaging in the tasks and interacting with the PST. During the interview, the interviewer prompted the PST to assess students' general responses to their instruction, and then to assess one or two particular students' responses. Some examples of questions were, "What was your goal for this lesson?," "How did the students of this class do with the content?," "Did you notice any difficulty that students were having? Where do you think those difficulties came from?," and "What would you do if you taught this lesson again?" The analyses of interview transcripts helped us to triangulate the patterns that emerged from the written teaching report (Denzin, 1978; Denzin & Lincoln, 2005; Merriam, 2009). In addition, the interviewer asked about PSTs' working relationships with their mentor teachers, specifically in designing their assessment tasks. This interview provided insights into PSTs' experiences in a local school context, and in particular their interactions with mentor teachers.

Lastly, we conducted individual *interviews with the two course instructors and 12 mentor teachers*. These were semistructured, hour-long interviews conducted near the end of the academic year. Similar to the interviews with PSTs, the researcher showed segments of PST teaching videos (the same ones used in the PST interviews) and asked similar questions. These comparison interviews were analyzed to understand the interviewees' relationships with their PSTs as well as the ways in which these significant professionals, who had regularly interacted with the PSTs, framed the goals of science learning. Some examples of questions were, "How do you usually teach this topic and why?," "What do you think the PST is trying to accomplish and how do you think about it?," "What do you like or dislike about PSTs' approach and why?," and "How did you work with your PST when they planned their instruction and after their instruction?"

Data Analysis

At the first stage, we analyzed PSTs' engagement in each of the assessment activities in response to RQ 1 to 3. Guided by the responsiveness framework and given the nature of assessment activities (see Figures 1 and 2), we coded PSTs' assessment tasks, interpretation, and reflection as documented in the 32 teaching episodes focusing on their attention and responsiveness. Next, cross-analyses were conducted across the 32 teaching episodes and 14 PSTs to examine (a) under which conditions PSTs successfully responded to student thinking and (b) what made PSTs more or less successful in responding to students' thinking. The following describes the details of the processes.

Phase I: Coding PSTs' Assessments, Interpretation, and Responses.

Assessment Tasks: Epistemic Frame Built in the Design (RQ 1). We analyzed the nature of the assessment tasks in each teaching episode holistically. We examined written descriptions of the goals, objectives, and designed/selected assessments, as well as the teaching artifacts (worksheets, slides), and teaching videos (if available) in each of the 32 teaching episodes. Our analysis was focused on explicating the nature of the opportunities PSTs provided for students to show both what they know and how they know through assessment tasks. The nature of these opportunities revealed how PSTs framed knowledge and learning of science as they designed or selected assessments in that instance.

Based on this analysis, we coded the assessment tasks as either productive or unproductive. Following recent influential documents (NGSS Lead States, 2013; NRC, 2007, 2012), we posited a *productive* epistemic frame as providing opportunities for students to engage in meaningful scientific practices for sense-making. We identified a total of four types of opportunities from our analysis of the assessments. Two types of opportunities were coded as a reflection of productive epistemic frames. In those opportunities the completion of the assessment tasks (items) required students to make connections between observable and unobservable (theoretical) elements of natural phenomena through intensive reasoning. Specifically, the assessment tasks prompted students to (a) construct scientific explanations or argumentation by reasoning through data, observation, or experiences and (b) use science ideas to account for observable phenomena. The assessment tasks coded as *unproductive* asked students to either (c) reproduce factual information or canonical scientific knowledge or (d) display skills or procedural knowledge (see the coding scheme and examples of assessment tasks in Table 2).

Interpreting Student Responses: Attending and Noticing (RQ 2). In this analysis, our goal was to investigate PSTs' interpretations of student responses, focusing on their attention and noticing. First, we identified documented student responses and the accompanying PSTs' accounts. Sometimes PSTs drew inferences about individual student responses; at other times, PSTs drew inferences about a pattern of collective student responses. We defined a PST's account, generated around either one individual response or a grouping of responses, as the unit of analysis. We identified a total of 154 units from 32 teaching episodes (4–5 units per one teaching episode). We analyzed each unit to identify what aspect(s) of student responses PSTs attended to. Building on the previous study (Kang & Anderson, 2012), these accounts were then categorized into three groups, attending to cognitive behavior, social behavior, and on-task behavior (see Table 3). When a PST attended to multiple student behaviors, all of that information was coded into a unit.

Next, we analyzed PSTs' noticing—how PSTs reasoned with the attended information to account for how and why students produced certain responses. During this process, PSTs *generated relationships* by connecting bits of attended information to one another. Eight initial codes emerged from the analysis of 154 units representing the relationships generated in this process. Specifically, PSTs related the attended student responses to (a) aspects of instructional design; (b) students' missing experiences that they didn't take into account, (c) failure to build upon students' prior experiences and knowledge; (d) home or family backgrounds, languages, and cultural resources (e.g., foster care, immigrant family); (e) social interactions in classrooms; (f) students' personalities, characteristics or working pattern; (g) students' attentiveness, attitude, or behaviors; and (h) problems of missing school knowledge prior to the year or misconceptions. We categorized these eight initial codes into three groups reflecting the ways in which PSTs *framed problems*. The three groups were (a) problems of instruction, (b) problems of learning environments, and (c) problems of students (see the final coding scheme in Table 3).

Suggesting Changes in Instruction: Productive vs. Unproductive Responses (RQ 3). In many teaching episodes, PSTs suggested changes collectively at the episode level. Therefore, the final analyses of PSTs' responses were coded at the teaching episode level. We coded the quality of PSTs' responsiveness to student thinking, as reflected in the suggested changes, as either productive or unproductive. Based on the literature on responsive teaching, instructional design (Coffey et al., 2012; Levin et al., 2009; Roseberry & Warren, 2008; Sohmer, Michaels, O'Connor, & Resnick, 2009) and the patterns that emerged from our data, we defined *productive responses* as specific changes in components of instruction

TABLE 2
Coding Scheme: Epistemic Frame Built in the Design of Assessments

Epistemic Frame	The Types of Opportunities	Examples of Assessment Tasks
Productive: Science learning targeted for assessments is framed as engaging students in meaningful scientific practices for active sense-making	a. Construct scientific explanations or argumentation by reasoning through data, observation, or experiences	*Susie #2, 10–11th grade chemistry, Phase change* Students, in teams, draw a picture of how ice melts in their hand and then answered: Q1. What is happening at a molecular level and where does the energy go? Q2. Why does the table surface feel "colder" than a wood surface? *Alisa #1, sixth-grade earth science, Soil* Students are introduced to a scenario and analyze the problem of a school garden that will not grow. At the end of the unit, students should describe what to do with the garden and make a proposal to the principal of the school. The examples of assessment questions are as follows: Q. Do you think your test results explain why the garden did not grow? Explain your ideas using the results from this activity as well as other activities in the unit. Q. Do you agree or disagree with the ideas of the students in the role-play? Explain why. Q. Describe a super soil that could be added to the school garden. Convince your classmates that you should use the soil by explaining how it will fix the garden problem
	b. Use science ideas to account for observable phenomena	*Monica #2, ninth-grade biology, Incomplete inheritance* Students made observation of the five different examples representing incomplete inheritance, came up with questions, and simulated the inheritance processes using a model. Several assessment questions appeared in the worksheet, including: Q1. What are some differences that you notice between this type of inheritance and the dominant/recessive pattern you learned about last week? Q2. Can you think of any examples of this type of inheritance that you've seen before? If so, what have you seen?

(*Continued*)

TABLE 2
Continued

Epistemic Frame	The Types of Opportunities	Examples of Assessment Tasks
Unproductive: Science learning targeted for assessments is framed as reproducing canonical scientific knowledge	c. Reproduce factual information or canonical scientific knowledge	*Lynn, 10–11th grade, Plant pathology* After students conducted plant pathology lab, five assessment questions were asked. Two examples were as follows: Q3. For each of the words listed below, decide whether it is a disease or symptom. Circle your answer. Q5. What is the disease triangle AND explain the three parts involve in this disease triangle. Make sure to explain relationship among the three parts.
	d. Display skills or procedural knowledge	*Kevin #2, ninth-grade physics, Two-dimensional motion vector components and addition/trigonometry* Q: If a block is sliding due north on a frozen (frictionless) lake and is then given a quick push to the east what happens to the blocks speed in the north/south direction? In the east/west direction?

TABLE 3
Coding Scheme: Interpreting Student Responses and Suggesting Changes

	Codes
Attending	*Cognitive behavior* (a) What students know and do not know accurately (i.e., attending to the substance of student ideas or thinking, such as partial understanding, alternative ideas, and misconception) (b) What students know and do not know broadly or inaccurately *Social behavior* (c) Social interaction in classroom (e.g., group dynamics) *On-task behavior* (d) Correctness of student responses (i.e., checking out whether student responses are correct or not) (e) Completeness and following directions
Noticing (generating relationships and framing problems)	*Problem of instructional design* (a) One or the other components of instruction (b) Students' missing experiences (c) Students' prior experience/background knowledge *Problem of leaning environment* (d) Home backgrounds, language, family, and cultural resources (e.g., foster care, immigrant family) (e) Social interactions in classroom *Problem of students* (f) Students' personalities, characteristics, or working pattern (g) Students' attentiveness, attitude, or behaviors (h) Problem of missing school knowledge prior to the year or misconception
Suggesting changes in instruction (responding)	*Productive responses*: Specific AND directly address the manifested students' difficulties (a) Modifying one or more components of instructional practices ▪ Task design (e.g., additional observation, complementary experiment) ▪ Talk patterns (e.g., asking different questions that draw students' attention to important information) ▪ Providing additional scaffolding (e.g., being explicit, modeling the processes, using sentence stems, changing participation structures) *Unproductive responses*: Generic and/or NOT address the manifested students' difficulties (b) Changing generic formats of instruction (e.g., lab vs. lecture) (c) General strategies to increase engagement (e.g., helping them to see the value of learning about it) (d) Generic strategies (e.g., more hands-on activity, making it relevant, spend more time on it) (e) Little or no changes

(e.g., task design, talk, tools, or scaffolds) that likely create better conditions to promote students' scientific thinking.

When a PST's responses were coded as *productive*, the changes suggested by the PST were specific and relevant—meaning they directly addressed the students' difficulties. For example, when a teacher sees that some students fail to notice important patterns across observations, a productive response might involve revising or adding questions to draw students' attention to the important patterns (changes in talk moves). In contrast, *unproductive responses* do not generate better conditions for student learning. For example, suggestions like "re-teach" or "do more labs" in a generic sense are unlikely to address students' particular difficulties. Typically, an *unproductive response* was generic or irrelevant to the manifested student ideas. Five codes emerged through the initial analysis, of which four were categorized as unproductive responses (see Table 3). Each of the 154 units was coded using the five subcodes.

Phase II: Cross-Episode Analyses—Examining PSTs' Engagement in Assessment Activities in Contexts (RQ 4). We examined the relationship between assessment task design, interpretation, and PSTs' responsiveness within and across the 32 teaching episodes to identify the conditions under which PSTs productively attended and responded to students' scientific thinking. We identified 12 teaching episodes that provided evidence of PSTs' productive responsiveness. With these 12 episodes, we traced back how PSTs' responsiveness was related to the patterns of assessment design and PSTs' interpretations. This analysis suggested some strong relationships between PSTs' productive responsiveness and the epistemic frame of the assessment tasks. Accordingly, we selected 18 teaching episodes that began with a productive epistemic frame, compared and contrasted PSTs' patterns of interpretations, and analyzed how those patterns were related to PSTs' productive or unproductive responsiveness. Finally, we examined the patterns within and across the remaining 14 teaching episodes that began with an unproductive epistemic frame, focusing on the PSTs' responsiveness.

Next, we shifted the grain size of the analysis from teaching episodes to the individual cases of the 14 PSTs. We categorized the 14 cases into three groups based on their responsiveness to student thinking. Employing the constant comparison method (Glaser & Strauss, 1967), we identified and compared the patterns within individual cases in each group, as well as across groups. We paid attention to other contextual information provided through interviews to understand how PSTs' assessment practices were shaped through their interactions with mentor teachers and teacher educators within contexts. Interview transcripts were analyzed focusing on the following questions: (a) how did PSTs interact with people at both the school and the program when they designed their assessments, and what resources were used by PSTs? (b) how were PSTs' interpretations of their assessments and student responses similar or different from mentor teacher's and course instructors'? (c) what were the contextual affordances or constraints for PSTs' productive responsiveness to student thinking? From these analyses, we intended to theorize PSTs' learning of responsive teaching in relation to the pedagogical approach of the program.

Subjectivity

There is a potential for bias in the process of collecting and interpreting data due to the relationships of the authors with the participants. The authors were either course instructors or field supervisors of five PSTs. On the other hand, the relationships provided some deeper insights into the nature of local school contexts, professional relationships and

interactions with their peers, mentors, and instructors, and their daily instructional practices in classrooms. We addressed this issue of potential bias with multiple layers of triangulation (Denzin, 1978) and the careful design of data collection. None of the authors were involved in the interviews with the participants who were under their supervision. Multiple sources and types of data collected through multiple methods were used to increase credibility. The generated coding scheme and interpretation were discussed and debated at weekly meetings for a total of two years. The codings, interpretations, and pictorial models were also presented and discussed with three other science education faculty members through a bi-weekly instructor meeting for about six months at the early stage of data analysis. In addition, a formal interrater reliability check was conducted with a doctoral student in science education program. We coded a sample of data together and revised the coding scheme iteratively until we reached up to 80% consistency.

FINDINGS

We present our findings in four sections, reflecting what we learned about the assessment tasks themselves and the ways in which PSTs interpreted and responded to students. The last section presents three illustrative cases to show how PSTs' attention and responsiveness to student thinking via these assessment activities were shaped in contexts.

Assessment Tasks

About a half of the teaching episodes used the productively framed design of assessment tasks that provided opportunities for students to reveal their sense-making (n = 18 out of 32, 56.3%). Among these 18 teaching episodes, two-thirds of the episodes provided evidence of PSTs' productive responsiveness to student thinking (12 out of 18, 66.7%; see Figure 3). There was no incidence of PSTs showing productive responsiveness to student thinking when unproductively framed assessments were used. About half of the productively framed episodes were observed from interns and the other half from seniors (8 from interns, 10 from seniors).

Ten out of 18 productively framed assessments were either designed or significantly modified by the PSTs with help from the program (e.g., using the provided lesson design

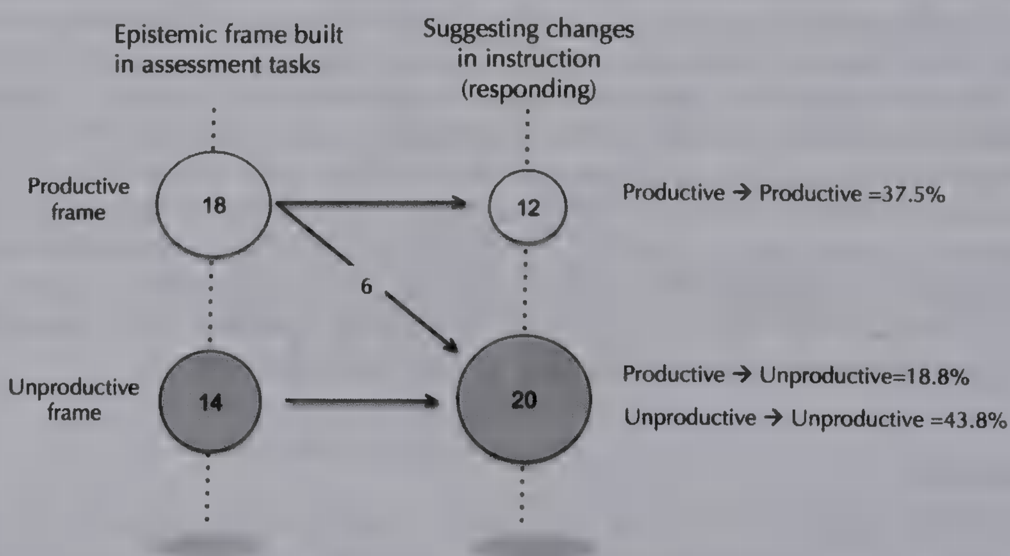


Figure 3. Epistemic frame built into the design of assessment tasks and teachers' responsiveness to student thinking.

TABLE 4
The results of analysis of 32 teaching episodes (The total number of the analyzed units = 154)

	PSTs	Teaching Episodes	Epistemic Frame Built in Assessment Tasks (D: Designed by PSTs, S: Selected)	No. of Units	Interpreting Student Responses: Attending to	Interpreting Student Responses: Framing Problems in Relation to	Suggesting Changes in Instruction
Student Thinking group	Leslie	#1 (Senior)	Productive (D)	3	Cognitive (1) Task behavior (2)	Instructional & student	Productive
		#2 (Senior)	Productive (D)	3	Cognitive behavior (3)	Instructional	Productive
		#3 (Intern)	Productive (D)	6*	Cognitive (6) Social behavior (1)	Instructional & student	Productive
		#4 (Intern)	Productive (D)	5	Cognitive behavior (5)	Instructional	Productive
	Alisa	#1 (Intern)	Productive (S)	8*	Cognitive (7) Social (1) Task behavior (4)	Instructional, student, environment	Productive
		#2 (Intern)	Productive (S)	7	Cognitive (6) Task behavior (1)	Instructional, student	Productive
	Scott	#1 (Senior)	Productive (D)	3	Cognitive behavior (3)	Instructional	Productive
		#2 (Senior)	Productive (D)	6	Cognitive behavior (4) Task behavior (2)	Instructional	Productive
Conditional group	Lori	#1 (Senior)	Unproductive (D)	3*	Cognitive (3) Task behavior (2)	Students	Unproductive
		#2 (Senior)	Productive (D)	10	Cognitive behavior (10)	Instruction	Productive
	Monica	#1 (Intern)	Unproductive (D)	3*	Cognitive behavior (3) Social (2)	Students	Unproductive
		#2 (Intern)	Productive (D)	6	Cognitive (3) Social (1) Task behavior (2)	Instruction	Productive
	Susie	#1 (Intern)	Unproductive (S)	3*	Cognitive (1) Task behavior (3)	Students	Unproductive
		#2 (Intern)	Productive (D)	6	Cognitive behavior (6)	Instruction	Productive
	Mary	#1 (Senior)	Productive (S)	3	Cognitive (2) Task behavior (1)	Students	Unproductive
		#2 (Senior)	Productive (S)	2	Cognitive behavior (2)	Instruction	Productive

(*Continued*)

TABLE 4
Continued

	PSTs	Teaching Episodes	Epistemic Frame Built in Assessment Tasks (D: Designed by PSTs, S: Selected)	No. of Units	Interpreting Student Responses: Attending to	Interpreting Student Responses: Framing Problems in Relation to	Suggesting Changes in Instruction
Responding to other concerns group	Shannon	#1 (Senior)	Productive (S)	3*	Task behavior (6)	Students	Unproductive
		#2 (Senior)	Productive (S)	2*	Cognitive (2) Task behavior (2)	Students	Unproductive
		#3 (Intern)	Unproductive (D)	3	Cognitive (1) Task behavior (2)	No connection	Unproductive
		#4 (Intern)	Unproductive (D)	4	Cognitive (1) Task behavior (3)	No connection	Unproductive
	Kevin	#1 (Senior)	Productive (S)	4	Cognitive (3) Task behavior (1)	No connection	Unproductive
		#2 (Senior)	Unproductive (S)	4	Task behavior (4)	No connection	Unproductive
	Stella	#1 (Senior)	Unproductive (D)	3*	Task behavior (6)	Students	Unproductive
		#2 (Senior)	Unproductive (D)	15	Cognitive (1) Task behavior (14)	Students	Unproductive
	Lynn	#1 (Senior)	Unproductive (D)	3	Task behavior (3)	No connection	Unproductive
		#2 (Senior)	Unproductive (D)	3	Cognitive behavior (1) Task behavior (2)	Students	Unproductive
	Adam	#1 (Intern)	Unproductive (D)	6	Task behavior (4) No highlighting (2)	Students	Unproductive
		#2 (Intern)	Productive (S)	4	Cognitive (1) Task (2) No highlighting (1)	Students	Unproductive
	Teresa	#1 (Intern)	Unproductive (D)	6	Task behavior (6)	Students	Unproductive
		#2 (Intern)	Productive (D)	6	Cognitive (2) Task behavior (4)	Students	Unproductive
	David	#1 (Intern)	Unproductive (D)	6	Task behavior (6)	Students	Unproductive
		#2 (Intern)	Unproductive (D)	5	Cognitive (2) Task behavior (3)	Students	Unproductive

*This unit includes multiple codings, therefore the sum of attended elements is bigger than the number of units.

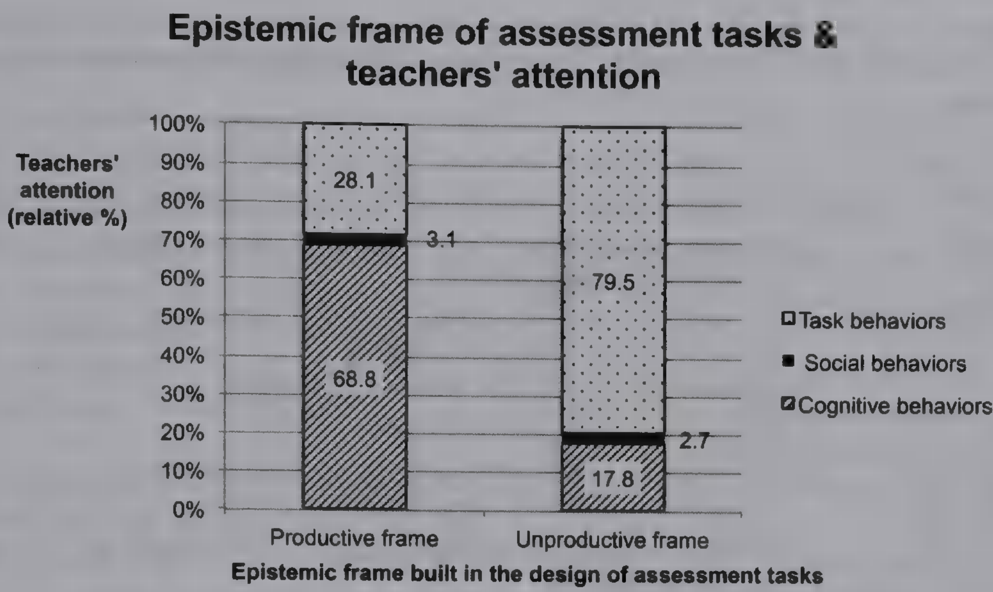


Figure 4. Epistemic frame of assessment tasks and teachers' attention.

framework and templates, using assessment tasks from the course websites with a slight modification, or codesigning the tasks with course instructors). The assessment tasks of the other eight teaching episodes came directly from mentor teachers or some commercial curricula with little modification. Notably, nine out of 12 teaching episodes that provided evidence of PSTs' productive responsiveness to student thinking were either designed or significantly modified by PSTs. Two of the remaining three were from curricula developed by university-based research groups.

Interpretation of Student Responses

Recall that we defined a "unit" of analysis as a set of accounts generated by a PST around a single or collective student response. Using this definition, we identified 154 units across the 32 teaching episodes. However, the total number of information units highlighted by PSTs was 169, because many PSTs attended to multiple pieces of information while simultaneously interpreting student responses (see Table 4). Overall, PSTs attended to on-task behavior most frequently ($n = 85$ out of 169, 50.3%), followed by cognitive behavior ($n = 79$ out of 169, 46.7%) and social behavior ($n = 5$ out of 169, 3.0%).

It appeared that PSTs' attention was strongly associated with the nature of assessment tasks. As shown in Figure 4, when productively framed tasks were used, PSTs attended to students' cognitive behavior about four times more than when unproductively framed assessment tasks were used (68.8% vs. 17.8%; see Figure 4).

PSTs linked the attended information to various aspects relevant to students' learning. In this process, about 70% of the attended student responses were connected to a problem with students, such as missing school knowledge, students' misconceptions, attentiveness, attitudes, personalities and task behaviors, etc. (69.8%). About one fourth of the attended information was attributed to a problem with instructional design (24.7%). In a few units, PSTs connected student responses to aspects of the context, such as personal, home, family backgrounds or social relationships, and interactions in the classroom (4.9%; see the detail in Table 5).

The ways in which PSTs made sense of student responses was strongly related to both the frame of the assessment task design and things that PSTs attended to. When

TABLE 5
Interpreting Student Responses: Connecting Attended Information to Some Problems

Problems Relevant to	Subcodes	%
Instructional design (24.7%)	• Problem of instructional approach (e.g., did not model it, did not provide scaffolds)	18.7
	• Did not know students do not have that experience	3.8
	• Did not consider students' prior experiences/background knowledge enough	2.2
Learning environment (4.9%)	• Social interactions in classroom	3.3
	• Home backgrounds, language, family, and cultural resources (e.g., foster care, immigrant family)	1.6
Student characteristics (69.8%)	• Students missed school knowledge prior to the year or had misconception	46.2
	• Students' attentiveness, attitude, behaviors, learning disabilities	20.9
	• Students' personalities, characteristics or working pattern	2.7
	• No connection	.5

productively framed assessments were used, PSTs connected students' responses to problems of instruction eight times more frequently (40.2% vs. 5.2%). In contrast, when unproductively framed assessment tasks were used, 92.2% of the attended information was connected to problems with students. Under the condition of attending to cognitive behavior, PSTs connected the attended student responses to problems of instruction about four times more than the condition of attending to on-task behavior (39.8% vs. 10.1%). In contrast, under the condition of attending to students' on-task behaviors, 89.8% of the attended information was linked to some problems with students.

Suggesting Changes in Instruction: Productive Versus Unproductive Responses. We analyzed the 12 teaching episodes that provided evidence of PSTs' productive responsiveness to student thinking. This analysis showed that both what PSTs attended to and the ways in which they framed problems while interpreting student responses played important roles in shaping their responsiveness. When PSTs attended to students' cognitive behavior, the likelihood of suggesting productive responses to student thinking was 73.4%. In contrast, it was only 10.6% if PSTs attended to on-task behaviors. When PSTs linked the attended information to problems relevant to instruction, they were about three times more likely to make suggestions for productive instructional modification than when they linked the information to problems with students (77.7% vs. 27.8%).

PSTs Attention and Responsiveness to Student Thinking via Assessment Activities in Contexts

The cross-analyses of 14 PSTs' cases revealed three distinctive patterns in the ways in which PSTs attended and responded to student thinking upon their engagement in assessment activities: (a) consistently attending and responding to student thinking (three PSTs); (b) conditionally attending and responding to student thinking (four PSTs); and (c)

responding to other concerns (seven PSTs) (see Table 4). In the following, we present an illustrative case from each of the three groups.

The Case of Leslie: Consistently Attending and Responding to Student Thinking With Use of Productively Framed Assessment Tasks

Leslie consistently paid great attention to students' ideas and thinking, and either suggested or adapted her teaching practices to address students' difficulties instructionally across all four teaching episodes in both her senior and internship years. In her senior year, Leslie was placed in a ninth-grade biology class at an urban, high-needs school that had a high percentage of ethnically, racially, and linguistically diverse students from low-income families. Leslie's mentor teacher, Mrs. F, had worked with the program for seven years. Mrs. F's instruction was fairly traditional—some combination of worksheets and hands-on activities, but she was supportive and open to new ideas. During the interview, Mrs. F described Leslie as an "active," "strong," and "confident" candidate who actively searched for useful resources and asked a lot of questions. Mrs. F. said, "[Leslie] would come to the discussions that she would have with some suggestions or ideas already prepared." The course instructor, Dr. G, described Leslie as "one of our top candidates" who was "really attentive to [the assignment] templates." Dr. G commented, "If we made comments, [she was] coming and asking questions about comment, even revising her unit or lesson plans, even if technically she didn't need to revise it. She followed the templates that we laid out very carefully."

Designing Assessment Tasks: Framing Science as Engagement in Meaningful Scientific Practices for Sense-Making. Leslie was assigned by her mentor teacher to teach the topic of chromosomes and the structure and function of DNA (Leslie's teaching episode #2 in Table 4). This lesson was anchored in one puzzling question, "If the DNA from a human cell was stretched out, it would be over 6ft tall! How do you think that a cell fits it all in?" Leslie began the lesson by asking students to think about the best way to pack a lot of clothes into a small suitcase. After eliciting students' ideas like "rolling" and "folding" through discussion, Leslie provided information about the process of folding DNA into chromosome with various pictures. Students labeled the pictures individually or in a group as walking through the worksheet. Students then built models of two nucleosomes with candy. Using this model, they explained the process of folding from the DNA string to chromosomes.

Leslie designed four assessment questions. First, she asked students to (a) "Label the picture. Be sure to use the following words: Histone (protein), DNA double helix, Chromatin, and Chromosome." Then, students had to answer three open-ended questions: (b) "Explain how DNA goes from a very long thin molecule to the thick structure of condensed chromosomes?" (c) "What are two reasons for folding DNA into chromosomes?" (d) "Next week we will be learning about when cells spilt in two. Why might chromosomes be important in the splitting process?" The design of this assessment provided opportunities for students to engage in modeling, and reason with both the information and the model to solve the puzzling question. The epistemic frame built in the design of this assessment task was coded as "productive" because science learning was framed as engagement in meaningful scientific practices for sense-making.

Interpreting Student Responses: Attending to Student Thinking and Framing the Problem in Relation to Instructional Design. Students produced various interesting

responses to this assessment task in both drawing and written form. Leslie attended to the details of student ideas and thinking, and identified their difficulties in sense-making (see the coding results of teaching episode #2 in Table 4). She then framed the problems by connecting them back to her instructional approach (problem of instructional design). For example, Leslie stated, "About half of my students thought that DNA turns into histones. In addition to the confusion around histones, I noticed that about 10% of my students thought chromatin and chromosomes were structures that DNA wrapped around. My students realized that chromosomes result from a series of folding and rolling steps, but they were confusing which step was which." Leslie then connected the student's confusion around histone with her unclear representation during the instruction. She stated, "There were a few things that I did during my lesson that might have confused them. For example, I showed a tangled mess of yarn and a ball of yarn to represent DNA being wrapped around a histone. The students never saw the bead the yarn was wrapped around, so they might have thought that the yarn itself represented the histone. If this was the case, then it is easy to see why so many of my students thought that DNA turned into a histone."

Suggesting Changes in Instruction: A Specific Modification of Task Design and Scaffolding. Leslie's responses were specific and tied to the difficulties that she identified in students. Leslie suggested two specific changes in components of instruction: "have students work together to draw a picture on the chalkboard or build a model that represents DNA at each stage of folding" (adapting task design) and "in the discussion, I would emphasize that a histone is a protein and the only structure that DNA wraps around" (purposeful talk moves). She stated,

> If I were teaching the next day, I would do a warm up question about DNA folding and then have a class discussion reviewing each structure. I would say something like "most of you seem to understand that strands of DNA are rolled and folded to make chromosomes, but many of you are confusing what each structure is composed of. Let's review this again." In addition to this, I might have my students work together to draw a picture on the chalkboard or build a model that represents DNA at each stage of folding. In the discussion I would emphasize that a histone is a protein and the only structure that DNA wraps around. I would also emphasize that chromatin and chromosomes include the DNA in their structure.

Across all four teaching episodes taught in her senior and internship years, Leslie continuously used assessment tasks that provided opportunities for students to engage in meaningful scientific practices and asked questions that generated complex and long responses. Leslie's reflection included specific ideas for changing components of instruction that were directly related to the manifested student difficulties.

The Case of David: Attending to On-Task Behaviors and Addressing Students' Wrong or Incomplete Responses With Generic Motivational Strategies. David was one of the seven PSTs in our "responding to other concerns" group. He was one of three PSTs who continuously used assessment tasks that were coded as *unproductive*. During his internship year, David worked with an exemplary mentor, Mrs. M, in an academically oriented high school near the university. Mrs. M was a veteran teacher who graduated from the same program 12 years ago. She was a highly regarded mentor teacher who provided PSTs with a good model of science teaching. During the interview, Mrs. M expressed her frustration in working with David throughout the internship year. Mrs. M said, "The planning portion, I didn't really feel like it was a strong relationship" and "Just in terms of behavior, he was

very closed. In terms of what you would see, his physical stance was with his arms crossed. Kind of staying at a further distance from me, and just saying, 'Okay, okay, okay,' and not really asking the 'why' part.' It was very short, like, 'I've heard you' and that was it." The course instructor, Dr. R described David's engagement in her methods course using the words like, "so flippant about things," "he is a joker," and "confident about his ability." Dr. R commented, "I think [David] thought he could be a fine teacher. I don't think he thought the program would do him a whole lot of good, you know. I think that he thought of himself as somebody who was a pretty good teacher coming in." Ironically, during the interview David commented that his instruction was pretty similar to Mrs. M and there was no big difference except "she has more stories to tell" due to her experiences.

Designing Assessment Tasks: Framing Science as Reproducing Canonical Scientific Knowledge. This teaching episode was focused on the human digestive system in ninth-grade biology (David's teaching episode #1 in Table 4). This unit was anchored on the phenomenon of a man eating a hamburger—What happens to a hamburger as it goes through the digestive system? While planning this unit, David selected following two questions as his assessments: (a) "Label a human body silhouette picture with the organs on it" and (b) "name two organs that digested a certain macromolecule." The key objective of this unit as identified by David was to "identify the correct location and relative size of the organs in each body system, and explain each organ's general function within its specific system." This assessment was coded as an "unproductive frame" because it only showed whether students were capable of reproducing facts or known scientific ideas.

Interpreting Student Responses: Attending to Task Behaviors and Connecting it to the Problem of Students. While analyzing student work, David primarily attended to correctness of student responses and students' on-task behavior (e.g., completing vs. not completing the work). For example, David narrated Pat's (one of his low-achieving students) responses as follows:

> Pat got all but three of them correct and for some reason she still labeled the bladder as part of the system. I am a little confused here because we haven't even talked about the urinary system yet and I made it a point to say that urine was not involved with the digestive system and that it only creates solid waste. This student is definitely more talkative in class, so there is a chance that she may have missed this. The only other organ she missed was the gall bladder which we didn't spend much time on other than just saying it was up under the liver and it stores the bile created by the liver.

David connected Pat's incorrect answer to her attentiveness to the instruction. Through the analysis of student responses, the main problem was framed as students' attentiveness, motivation, and behavioral issues (i.e., problems of students).

Suggesting Changes in Instruction. To improve the instruction for this topic, David suggested adding "some sort of activity or project."

> I really liked my lesson sequence for this unit and think that it went well . . . the only thing I would change is that for the final assessment of this lesson sequence I might have them do some sort of activity or project with it where they maybe make a working model of the digestive system or something cool like that.

Despite the potential for increasing students' attentiveness to the task with this hands-on activity, the instructional modification of "doing some sort of activity" was less likely to address Pat's difficulties in understanding why the bladder is not part of the digestive system. This kind of positive and overall satisfactory comment about the lessons was another salient pattern among the PSTs in this *responding to other concerns* group. Typically the suggested ideas for adapting instruction were generic or too general, and were not related to the students' manifested intellectual difficulties.

The Case of Monica: Attending to Students' Social Behaviors and Responding With General Strategies of Action. Monica was one of three PSTs who was successful in providing evidence of responsiveness in one of two teaching episodes—the one she taught at the later stage of the year. There were noticeable differences between the two teaching episodes (see Monica's #1 and #2 in Table 4). One difference was her changes in the design of assessment tasks to provide new forms of opportunities (from an unproductive to a productive frame). The other was Monica's increasing attentiveness to students' scientific thinking with the use of the productively framed assessment tasks. But across the two episodes, Monica consistently framed problems in relation to her instructional design.

Throughout her time in the program, Monica had struggled to teach science "differently," meaning "not lecturing all the time." She said, "Up until college, none of my classes really focused on inquiry or application as far as I can remember. Like, as a student, you're not really thinking about that." During the interview, Monica commented, "I get tired of lecturing and [students] don't like lecturing. My first lesson I taught to seniors, I lectured the entire hour. Their teacher must have threatened them because they were really good."

In her internship year, Monica worked with Ms. S who was fairly traditional in her instructional approach and less supportive of the program's approach. Monica aspired to create an "emotionally supportive but cognitively challenging" classroom learning community where "everyone can be successful." Monica's mentor, Ms. S said, "[Monica] was quite a researcher, so she was really good about going out and trying to find something new." Monica actively interacted with the methods course instructors, field supervisors, and her peers to receive feedback on her plans. Her course instructor, Dr. R said, "Monica asked a lot of questions both before she would plan, but also while she was teaching. She would sometimes write me an email, or when we saw each other in class, she really wanted to talk about some things."

Designing Assessment Tasks: From Knowing Canonical Science Ideas to Engaging in Sense-Making. The topic of her first teaching episode was osmosis and diffusion within a cell in a ninth-grade biology course. The assessment used in this episode consisted of five true/false questions, three short answer questions, and one multiple-choice question. One key assessment question that Monica "really hoped [students would] understand by the end of unit" was a multiple-choice question about the mechanism of osmosis. In the pre-assessment, students were asked to choose one correct prediction about the movement of water inside a U-shaped tube over time, from among four choices. In the end-of-unit assessment, Monica asked students to mark the status of the water inside a U-shape tube at equilibrium. Both assessment tasks were coded as "unproductive frame" because, by design, the assessments revealed whether students knew the canonical science ideas (osmosis) that were covered in the lecture.

In the other teaching episode (Monica #2 in Table 4), Monica taught a complex inheritance unit, following Mendelian genetics in ninth grade. During the sequence of activities, students made observations of the five pictures showing that offspring displayed a mixture

of parents' traits as opposed to displaying a single parent's trait—the five examples of complex inheritance from students' everyday lives. Students discussed their observations to formulate questions or hypotheses, and engaged in a modeling activity that illustrated incomplete or codominant inheritance patterns. The two key assessment questions that appeared on the worksheet were as follows: (a) What are some differences that you see between this type of inheritance and the dominant/recessive pattern you saw last week? (b) Can you think of any examples of this type of inheritance that you've seen before? If so, what have you seen? In addition, students had to go back to the questions that they formulated at the beginning of this unit and answer those questions based on what they learned from the modeling activities. This assessment was coded as productive because it provided opportunities for students to engage in scientific practices (pattern finding), and to use this idea to make sense of complex real-world phenomena.

Interpreting Student Responses: Attending to Learning Environments That Affect Students' Engagement and Framing the Problem in Relation to Learning Environments. The analyses across teaching episodes suggested that Monica was very attentive to students' personal backgrounds, language, and social relationships both in and outside her classroom. She drew upon that information to make sense of student responses. Monica's report, especially for the first teaching episode, included a variety of information about the three focus students. For example, Monica noted that Ken, her high achieving student, moved to this school from Japan a year ago, and had been struggling with scientific terminology despite his proficiency in English. Monica stated about Tom, who was a lower academic achiever, "Although my initial opinions were based on his standardized test scores, Tom is a very bright student who expresses his love of science—struggles to stay focused in social situations, but an active participant in class overall."

While making sense of students' responses produced from her assessment questions, Monica attended to students' cognitive and social behavior, and she connected the information to some aspects of the learning environment offered to the students. For example, in the post assessment about osmosis, all the students except Tom failed to generate a scientifically accurate prediction of the change of water level. Tom correctly drew the change of the water level, but showed that all of the particles moved to the same side, which was contradictory to his prediction about the changes of the water level. Monica noted, "Tom was the only one to make the water level higher on one side than on the other, but he also showed that all of the particles moved to the same side as well. I'm not sure if he thinks that when water moves, the particles must travel with it, or whether this was just a guess." Monica also connected Tom's responses to some personal factors outside of the classroom ("recently placed in foster care," "the number of class days he missed") and then problematized the general instructional approach (i.e., lab vs. lecture):

> Tom seemed to struggle a bit on this assessment. He was recently placed in foster care and I think the number of class days he missed drastically affected his results. In general, he seemed to struggle most on the application questions that not only used information from this unit but also required students to draw upon knowledge from prior units . . . In general, Tom seemed to excel on questions that he had the most experience with (labs, concepts we discussed frequently in class). The concepts that he learned only in lecture format however (active/bulk transport) he seemed unprepared to discuss.

In contrast, Monica's interpretations about student responses in the Genetics teaching episode was much more detailed and attentive to cognitive behavior relevant to scientific sense-making. For instance, Monica noticed students' difficulties in differentiating

observation from inference, and its consequential effect on their sense-making: "From what I observed, many of the students were making inferences rather than observations. For example, instead of stating that the flower contained both pink and white petals, students would say that both pink and white were dominant."

Suggesting Changes in Instruction: Modifying the Design of Tasks. In the osmosis teaching episode, the suggested ideas for adapting instruction were promising in principle, but not specifically tied to the students' manifested difficulties. Monica suggested a modification of her planning approach to "incorporate more student experiences initially." Despite the fact that Monica noticed all students except Tom responded to her "most important question" inaccurately, and even Tom, who drew the water level accurately, did not show his understanding about the mechanism of osmosis, she did not make any suggestions to address this problem.

In the genetics teaching episodes, Monica noted that students "did not do a great job of connecting real world examples to these patterns." She specifically pointed out students' confusion between two distinct patterns (codominance vs. incomplete dominance), and described the actual changes that she made on the following day (adding complementary tasks).

DISCUSSION

This study examined 14 PSTs' responsiveness to student thinking while engaged in assessment activities. Based on the findings, we offer two speculations about how teacher education programs might be able to better help PSTs attend and respond to student thinking. The first has to do with ensuring that PSTs have an opportunity to use more productive assessment tasks. The second has to do with when and where PSTs engage in interpreting student responses.

Using High-Quality Assessments Creates Opportunities for PSTs to Attend and Respond to Students' Scientific Thinking

Perhaps the most important finding from this study is that PSTs' productive responsiveness was *only* observed when they used productive assessment tasks—that is, tasks that appeared to provide opportunities for students to engage in scientific sense-making. In addition, the use of high-quality assessments was strongly related to PSTs' attention to students' cognitive behaviors and responsiveness to student thinking. We know that the design qualities of the assessments themselves significantly affected the quality of produced responses (Herman, 1992; Supovitz, 2012). Well-designed assessments make various ideas and student thinking visible (Furtak & Ruiz-Primo, 2008; Kang et al., 2014), which creates a condition for PSTs to notice students' ideas and ways of reasoning. In our data set, when productively framed assessments were used, student responses tended to be long and complex. In contrast, students mostly produced short and simple responses when unproductively framed assessment tasks were used. When unproductively framed assessments were used, PSTs mostly attended and responded to students' social or task behaviors.

While the program involved in this study sought to promote more productive assessments, PSTs used them in only 14 out of 32 teaching episodes (43.8%). We can only speculate on the reasons why high-quality assessments were not selected more frequently. It can be hypothesized that the variance has to do with disparities between program goals and

the philosophies of mentor teachers—what is sometimes called the "two-worlds pitfall" (Anagnostopoulos, Smith, & Basmadjian, 2007; Feiman-Nemser & Buchmann, 1985). The nature of student teaching is such that PSTs encounter multiple frames of science knowledge and learning as well as competing expectations between program and school, which might make it difficult for PSTs to select and use high-quality assessments provided by the program.

In our study, this hypothesis is partially supported by two cases in the *Conditional Group*, including Monica's. Both of these PSTs worked with mentor teachers who had strong ideas about the goals of science learning that contradicted those held by the program. Initially both of the PSTs mostly relied on their mentor teachers' curricula, including assessment tasks. Later they designed their own assessment tasks, while actively using tools supplied by the program and negotiating them locally as they worked with mentor teachers (see the case of Monica). However, the fact that only two out of 14 cases provided support to this hypothesis, and more importantly some cases like David provided counter evidence, seems to undermine this theory of the two worlds problem. Instead, we conjecture that the selection of high-quality assessments has something to do with complex social interactions among PSTs, course instructors, and school professionals in the process of planning (see Martin, Snow, & Franklin Torrez, 2011).

It is also possible that the relatively low percentages of the high-quality assessments have to do with PSTs' perspectives on the goals of science learning and knowledge, grounded in their past experiences—PSTs' initial, and apparently unchanged epistemic frames during the preparation period. Three of the cases in the *responding to other concerns* group support this hypothesis. These three PSTs commonly worked with exemplary mentor teachers who were highly regarded by the program receiving relatively consistent and coherent messages from both program and school. Despite the available assessment tasks from the program and the feedback provided by course instructors, in the end, high-quality assessments were not selected by these PSTs and the suggested ideas for modifying the design of assessments were not taken seriously across all the observed teaching episodes.

What made the PSTs less interested in changing their assessments? One hypothesis is that it might be because these PSTs thought of their assessments as good and productive. Like most teachers, PSTs selected items that they believed would inform them about students' success in achieving their instructional goals. Thus, PSTs who came with unproductive epistemic frames selected low-cognitive-level assessment items that were consistent with their frames. For example, David, who consistently used unproductively framed assessments, expressed satisfaction with his instruction and students' responses, which was contradictory to mentor and course instructor's strong concerns. By choosing unproductively framed assessment tasks to begin with, David created qualitatively different conditions to learn about student thinking.

In contrast, some PSTs like Leslie—who came in with a strong interest in students' ideas and thinking and framed teaching as working on students' ideas to begin with—actively interacted with course instructors and chose the high-quality assessment tasks that revealed rich information about students' ideas. Those PSTs created ample opportunities for themselves to attend and respond to student thinking through their active choice of assessment tasks. The assessment activity was designed to help PSTs learn about student thinking and how to respond to it. However, the analysis showed that *PSTs actively created different qualities of opportunities for their own learning with their choices of assessment tasks*. And over time, the PSTs like Leslie learned more about students' ideas and developed their capacity for responsive teaching.

This pattern of "the rich get richer, the poor get poorer" has also been suggested by other researchers. In a study about beginning teachers' collective inquiry about student work, for

example, Windschitl, Thompson, and Braaten (2011) found that beginning teachers who had complex and problematized views on teaching and learning, such as Leslie, derived the most benefit from the collegial inquiry into student work. In contrast, the teachers with simplistic and unproblematized views, such as David, did not show significant learning gains through the analysis of student work.

Lastly, the reason why high-quality assessment tasks were not used by PSTs might have to do with the intertwined nature of instruction, assessment, and standards. Teachers cannot assess something that they do not teach. In general, teachers do not teach the things that are not listed in the standards. Even though some PSTs aspired to use productively framed assessment tasks provided by the program, they might not be able to select those assessments if their instruction (or their mentor's) did not match the assessments. In her teaching report, Monica described her selection of assessment tasks as "at this point I am still using my mentor teacher's materials." Monica might not have been able to select assessments during the period when she primarily followed Ms. S's instructional approaches. David might not have seen a reason to change his assessment in part because he thought his assessment aligned with the standards.

In short, our analysis clearly suggests that the use of high-quality assessment tasks that visualize students' thinking are necessary to promote PSTs' responsiveness to student thinking. Despite this obvious condition, creating productive situations for PSTs in student teaching contexts is not a simple task for teacher educators. The analysis reveals the contentious nature of selecting and using assessment tasks at the intersection of multiple expectations from PSTs, school professionals, and university programs.

Helping PSTs to Reframe Problems by Deprivatizing Interpretation

The use of high-quality assessments was necessary, but not sufficient, to promote PSTs' abilities to attend and respond to students' scientific thinking. One-third of the teaching episodes that began with high-quality assessment tasks failed to yield evidence of PSTs' responsiveness to student thinking ($n = 6$ out of 18, 33.3%). Analysis of these six teaching episodes showed PSTs' difficulties in noticing and reasoning about student thinking even when they were provided rich information about students' ideas. Other researchers have documented similar findings (see Gearhart et al., 2006; Lyon, 2013; Maclellan, 2004; Morgan & Watson, 2002). They pointed out that novice teachers' underdeveloped knowledge about content, teaching and learning, assessment, facility, or skills about assessment can explain these difficulties. Building on this previous work, in this study we attended to PSTs' framing and ability to notice in accounting for the five PSTs' uneven success in responding to students' thinking under the condition of using high-quality assessment tasks.

We conjecture that some PSTs' failure to attend to students' cognitive behavior, even with use of high-quality assessment tasks, had to do with novice teachers' *frame shifting* abilities. For instance Shannon, who was partnered with Leslie and taught the same microscope lesson using identical teaching materials and assessment tasks to a different period in Mrs. F's classroom, only noticed students' task behaviors. Both Leslie and Shannon described their difficulties in managing the classroom in their reports as well as during the interviews. Leslie noticed and responded to various student ideas along with students' task behaviors. But Shannon only highlighted students' task behavior while analyzing student work (see Table 4). Classrooms are complex spaces, especially to novice teachers, where teachers must be able to simultaneously monitor multiple problems within a crowd of students (Kennedy, 2005). Expert science teachers who carry out effective instruction are capable of attending and responding to various issues moment by moment while *flexibly shifting their framing*—their sense of "what is going on here." We speculate that

novice teachers' inflexibility in shifting their framing has significant influence on their responsiveness.

Attending and responding to students' cognitive, social, and task behaviors are all legitimate and essential skills to successfully helping students to learn. To support novice teachers as they gradually but systematically develop this expert-like ability, teacher educators can reduce the complexity of the work of teaching by having PSTs focus on one aspect at a time (Lampert et al., 2013). Within the structure of this assessment activity, PSTs were expected to focus on students' ideas and thinking and learn how to respond to it. However, for some PSTs like Shannon and Mary, who seemed to be overwhelmed with management-related concerns in the early stages of their student teaching, engaging PSTs in this assessment activity did not seem to be sufficient to draw their attention to students' cognitive behavior and notice students' ideas and thinking. Alternatively, Mary's and Shannon's difficulties in attending to students' cognitive behaviors through the use of high-quality assessments might have been related to their underdeveloped knowledge and skills regarding content, teaching and learning, and assessment—the essential components for noticing (Sherin & van Es, 2005; van Es & Sherin, 2008). Their weak content knowledge and skills might have led them to attend to students' task behaviors, instead of cognitive behaviors. However, Mary succeeded in attending and responding to student thinking in her next teaching episode a month later. Assuming that Mary's knowledge and skills would not have grown to such a degree within a single month, this knowledge theory appears to be less convincing in accounting for novice teachers' failure of attending and responding to students' scientific thinking than our hypothesis of flexible framing shift when attending and interpreting student responses.

Some PSTs' failure to respond to students' thinking under the condition of using high-quality assessment tasks points to the problem of the privatized nature of the interpretation in the design of this learning opportunity. One important pattern that emerged from the six teaching episodes was the way in which PSTs framed the problem while interpreting student responses, and how it led to proposed instructional modifications. In most cases, the "breakdown" on the pathway from attending and responding to student thinking was observed when PSTs reasoned with student responses during interpretation. All five PSTs linked students' failure to provide expected responses to some kind of problem *with* the students. The most popular link was to "lack of prior knowledge" without unpacking what that meant in the instructional context. Notably, in those cases, the suggested instructional modifications were generic or unrelevant, reflecting PSTs' limited repertoires and strategies.

In contrast, the PSTs who consistently attended and responded to student thinking, including Leslie, actively used the language from the program as her pedagogical reasoning resource to account for the observed student responses. PSTs' responsiveness hinged upon their interpretation of students' responses, in particular, the ways in which they framed the problem during the process, but there was little evidence that this process was interrupted during their engagement with this assessment activity. Within the structure of the activity and despite the support given by the program, such as one session of collective inquiry into student work, the use of the alternative reasoning resources was largely dependent upon PSTs' personal choices.

Taken together, setting up opportunities for PSTs to learn about students' ideas and thinking and how to respond to it requires more than having PSTs use high-quality assessments. Both PSTs' frame shifting that results in noticing students' intellectual difficulties and their interpretation of student responses need to be carefully scaffolded; therefore, PSTs can reframe the problem by considering alternative reasoning resources beyond their own ideas.

CONCLUSIONS

Over the last several decades, the field of education research on student learning has made significant progress. This leads teacher educators to have strong hypotheses about student learning and essential instructional capacities that are necessary to support students' learning. In contrast, the research on teacher education is still at the adolescent stage (Grossman & McDonald, 2008). The field is only beginning to understand how teachers develop essential instructional capacities and design professional learning opportunities informed by empirical evidence. This study contributes to our knowledge base about preservice teachers' learning about one important instructional capacity—attending and responding to student thinking.

The findings point to two considerations in designing learning opportunities to enhance PSTs' responsiveness to student thinking: (a) the use of high-quality assessment tasks that make student thinking visible and (b) helping PSTs to reframe the problems by deprivatizing PSTs' interpretation about student responses. This study demonstrates that without using high-quality assessment tasks and without pressing PSTs to reframe problems in relation to their instructional design, it is unlikely that PSTs will attend and respond to student thinking.

Even though key components of effective learning environments are apparent, designing learning opportunities that create conditions for PSTs' to be responsive to student thinking during the preparation period is a complex task. Teacher educators must strategically work with a *system* in a way to address various kinds of contextual and institutional challenges in designing learning activities for PSTs. The program in this study, for example, later changed the structures of the activities based on the lessons from this research project. In the modified assessment activities, each PST was required to use at least three assessments—one from their mentor teacher, one recommended by the program, and a third one that the PST prefers.

In the field of mathematics education, some researchers suggest using a carefully chosen set of *instructional activities* to support novice teachers in advancing their practices toward ambitious goals of mathematics learning (Lampert & Graziani, 2009). Instead of depending on novice teachers' abilities to design and select assessment tasks, using preselected high-quality activities and assessment tasks as a common context for learning about teaching can be another way of addressing the problems. Recent professional development approaches that relocate the places of teacher learning into classrooms, such as Studio approaches or collective Lesson Study, can be promising models for deprivatizing the work of interpretation (see Lampert et al., 2013; McDonald, Kazemi, & Kavanagh, 2013; Teachers Development Group, 2010). This kind of design increases the space for teacher educators to step in and to provide alternative pedagogical reasoning resources in the moment; therefore, increasing the likelihood for PSTs to better respond to students' thinking while reframing the problem collectively and collaboratively.

We know that what teachers do in the classroom has a huge impact on students' learning. It is more influential for students' academic futures than any other in-school factor, including quality of curricula, type of school, length of school year, or peer influences (see Murnane & Steele, 2007; Rockoff, 2004; Sanders & Rivers, 1996). Preparing future science teachers who are capable of attending and responding to various students' needs is one imperative agenda. We can achieve this goal only when we, the teacher education community, cultivate responsive mechanisms while continuously revising our understanding of how, and under what conditions, PSTs productively learn from practices. This study sheds light on the complex processes of PSTs' learning in the context of a teacher preparation program, which provides a foundation for teacher educators' responsive practices.

We are grateful to the preservice teachers, mentor teachers, and science education faculty at the Michigan State University who made this research possible. Many thanks go out to Mary Kennedy, Elizabeth van Es, and Lynsey Gibbons who provided their thoughtful comments on this paper. This article benefitted from critical feedback from two anonymous reviewers and the editor of this journal.

REFERENCES

Anagnostopoulos, D., Smith, E. R., & Basmadjian, K. G. (2007). Bridging the university–School divide. Journal of Teacher Education, 58(2), 138–152.

Atkin, J. M., Coffey, J. E., Morthy, S., Sato, M., & Thibeault, M. (2005). Designing everyday assessment in the science classroom. New York, NY: Teacher College Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. Phi Delta Kappan, 86(1), 9–21.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139–148.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academies Press.

Calabrese Barton, A., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science. American Educational Research Journal, 50(1), 37–75.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2012). The missing disciplinary substance of formative assessment. Journal of Research in Science Teaching, 48(10), 1109–1136.

Denzin, N. (1978). The research act: A theoretical introduction to sociological methods (2nd ed.). New York, NY: McGraw-Hill.

Denzin, N., & Lincoln, Y. (2005). Introduction: The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), Handbook of qualitative research (Vol. 3, pp. 1–32). Thousand Oaks, CA: Sage.

Elby, A., Richards, J., Walkoe, J., Gupta, A., Russ, R. S., Luna, M. J., ... Sherin, M. G. (2014). Differing notions of responsive teaching across mathematics and science: Does the discipline matter? Paper presented at the 2014 International Society of the Learning Sciences, Boulder, CO.

Feiman-Nemser, S., & Buchmann, M. (1985). Pitfalls of experience in teacher preparation. Teachers College Record, 87(1), 53–65.

Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. Journal of Research in Science Teaching, 49(9), 1181–1210.

Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. Science Education, 92(5), 799–824.

Gay, G. (2000). Culturally responsive teaching: Theory, research, & practice. New York, NY: Teachers College Press.

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S., Schwab, C., Vendlinski, T., ... Bernhaum, D. J. (2006). Developing expertise with classroom assessment in K-12 science; Learning to interpret student work. Educational Assessment, 11(3 & 4), 237–263.

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine Publishing Company.

Goffman, E. (1974). Framing analysis: An essay on the organization of experience. New York, NY: Harper & Row.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. American Educational Research Journal, 45(1), 184–205.

Hammer, D., Goldberg, F., & Fargason, S. (2012). Responsive teaching and the beginnings of energy in a third grade classroom. Review of science, mathematics and ICT education, 6(1), 51–72.

Herman, J. L. (1992). What research tells us about good assessment. Educational Leadership, 49(8), 74–78.

Interstate New Teacher Assessment and Support Consortium Science Standards Drafting Committee. (2002). Modeling standards in science for beginning teacher licensing and development: A resource for state diaglogue. Washington, DC: Council of Chief State School Officers.

Kang, H., & Anderson, C. W. (2012). The mechanisms of secondary science teacher candidates' learning to teach. Paper presented at the NARST, Indianapolis, IN.

Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show what they know: The role of scaffolding in formative assessments. Science Education, 98(4), 549–742.

Kennedy, M. M. (1999). The role of pre-service teacher education. In L. Darling-Hammond & G. Sykes (Eds.), Teaching as the learning profession: Handbook of policy and practice (pp. 54–86). San Francisco, CA: Jossey Bass.

Kennedy, M. M. (2005). Inside Teaching: How Classroom Life undermines reform. Cambridge MA: Harvard University Press.
Kloser, M. (2014). Identifying a core set of science teaching practices: A Delphi expert panel approach. Journal of Research in Science Teaching, 51, 1185–1217.
Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., ... Crowe, K. (2013). Keeping It Complex: Using Rehearsals to Support Novice Teacher Learning of Ambitious Teaching. Journal of Teacher Education, 64(3), 226–243.
Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. The Elementary School Journal, 109(5), 491–509.
Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. Journal of Teacher Education, 60(2), 142–154.
Lortie, D. C. (1975). School teachers: A sociological study. Chicago: University of Chicago Press.
Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. Science Education, 97(3), 442–467.
Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. Teaching and Teacher Education, 20(5), 523–535.
Martin, S. D., Snow, J. L., & Franklin Torrez, C. A. (2011). Navigating the terrain of third space: Tensions with/in relationships in school-university partnerships. Journal of Teacher Education, 62(3), 299–311.
Maskiewicz, A. C., & Winters, V. A. (2012). Understanding the co-construction of inquiry practices: A case study of a responsive teaching environment. Journal of Research in Science Teaching, 49(4), 429–464.
McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. Journal of Teacher Education, 64(5), 378–386.
Merriam, S. B. (2009). Qualitative research: A guide to design and implementation. San Francisco, CA: Jossey-Bass.
Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. Review of Educational Research, 65(2), 93–127.
Metz, K. E. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. Cognition and Instruction, 22(2), 219–290.
Minsky, M. (1985). The society of mind. New York, NY: Simon and Schuster.
Morgan, C., & Watson, A. (2002). The interpretative nature of teachers' assessment of students' mathematics: Issues for equity. Journal for Research in Mathematics Education, 33(2), 78–110.
Murnane, R. J., & Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. The Future of Children, 17(1), 15–43.
Nasir, N. i. S., Rosebery, A. S., Warren, B., & Lee, C. (2006). Learning as a cultural process: Achieving equity through diversity. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 489–504). New York, NY: Cambridge University Press.
NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington, DC: The National Academies Press.
NRC. (2005). How students learn science in the classroom. Washington, DC: National Academies Press.
NRC. (2007). Taking science to school: Learning and teaching science in grade K-8. Washington, DC: National Academies Press.
NRC. (2010). Preparing teachers: Building evidence for sound policy. Washingto n, DC: The National Academies Press.
NRC. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core Ideas. Washington, DC: National Academies Press.
Otero, V. K. (2006). Moving beyond the "get it or don't" conception of formative assessment. Journal of Teacher Education, 57(3), 247–255.
Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. Journal of Research in Science Teaching, 45(4), 497–523.
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. The American Economic Review, 94(2), 247–252.
Roseberry, A., & Warren, B. (2008). Teaching science as English language learners: Building on students' strengths. Arlington, VA: NSTA Press.
Rosebery, A. S., & Puttick, G. M. (1998). Teacher professional development as situated sense-making: A case study in science education. Science Education, 82, 649–677.
Ross, R. (1975). Ellipsis and the structure of expectation. San Jose State Occational Papaers in Linguistics, 1, 183–191.
Russ, R. S., & Luna, M. J. (2013). Inferring teacher epistemological framing from local patterns in teacher noticing. Journal of Research in Science Teaching, 50(3), 284–314.

Sadler, D. R. (1998). Formative assessment: Revising the territory. Assessment in Education: Principles, Policy & Practice, 5, 77–84.
Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
Sawyer, K. R. (2006). The new science of learning. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 1–16). New York, NY: Cambridge University Press.
Schank, R. (1990). Tell me a story: A new look at real and artificial memory. New York, NY: Scribner.
Schön, D. A. (1983). The reflective practitioner: How professionals think in action. New York: HarperCollins College Publishers.
Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011). Mathematics teacher noticing: Seeing through teachers' eyes. New York, NY: Taylor & Francis US.
Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. Journal of Technology and Teacher Education, 13(3), 475–491.
Sohmer, R., Michaels, S., O'Connor, K. M., & Resnick, L. B. (2009). Guided construction of knowledge in the classroom: The troika of talk, tasks and tools. In B. Schwarz, T. Dreyfus & R. Hershkowitz (Eds.), Transformation of knowledge through classroom interaction. New York, NY: Routledge.
Stake, R. (2004). Qualitative case studies. In N. Denzin & Y. Lincoln (Eds.), The Sage handbook of qualitative research, (3rd ed.) (pp. 443–466). Thousand Oaks, CA: Sage.
Supovitz, J. (2012). Getting at student understanding–The key to teachers' use of test data. Teachers College Record, 114, 1–29.
Tannen, D. (1993). Framing in discourse. New York, NY: Oxford University Press.
Teachers Development Group (Producer). (2010). Mathematics Studio Program: Transforming a school's culture of mathematics professional learning. Retrieved from https://http://www.teachersdg.org/default.htm
Thompson, J., Hagenah, S., Kang, H., Stroupe, D., Braaten, M., Colley, C., & Windschitl, M. (2016). Rigor and responsiveness in classroom activity. Teachers College Record, 118(7).
van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. Teaching and Teacher Education, 24(2), 244–276.
Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: Will it work? In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 53–82). New York, NY: Routledge.
Windschitl, M. (2005). Guest editorial: The future of science teacher preparation in America: Where is the evidence to inform program design and guide responsible policy decisions? Science Education, 89(4), 525–534.
Windschitl, M., Thompson, J., & Braaten, M. (2011). Ambitious pedagogy by novice teachers: Who benefits from tool-supported collaborative inquiry into practice and why? Teachers College Record, 113(7), 1311–1360.
Yin, R. K. (1989). Case study research: Design and methods (2nd ed.). Newbury Park, CA: Sage.

Science
Education

Science Teaching Reform Through Professional Development: Teachers' Use of a Scientific Classroom Discourse Community Model

ELIZABETH B. LEWIS,[1] DALE R. BAKER,[2] BRANDON A. HELDING[3]
[1]*Department of Teaching, Learning, & Teacher Education, College of Education & Human Science, University of Nebraska–Lincoln, Lincoln, NE 68588, USA;* [2]*College of Education-Secondary Education, Arizona State University, Tempe, AZ 95287, USA;* [3]*Boulder Language Technologies, Boulder, CO 80301, USA*

Received 2 January 2014; accepted 18 February 2015
DOI 10.1002/sce.21170
Published online 11 May 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This report outlines a 2-year investigation into how secondary science teachers used professional development (PD) to build scientific classroom discourse communities (SCDCs). Observation data, teacher, student, and school demographic information were used to build a hierarchical linear model. The length of time that teachers received PD was the exclusive predictor of change over time, whereas a schools' percentage of low socioeconomic students predicted how PD concepts was initially implemented. Prior to PD teachers expressed a desire to increase opportunities for students to engage in SCDCs, but found some aspects more challenging than others to implement. Generally, there were three categories of the teachers' frequency of use of SCDC strategies: (a) *most observed* that required teachers to change their own communication, classroom management, and direct instruction; (b) *occasionally observed* that provided opportunities for greater oral and written discourse to facilitate students' meaning making of science; and (c) *least observed* that encouraged students' executive control of their learning and teachers' use of formative assessment in response to students' diverse learning needs. Teachers identified administrative support, PD strategies, and teacher collaboration as supports for implementation. However, they rated students' science knowledge, diverse language skills, and discourse abilities as the greatest barriers to implementing a SCDC. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**896–931, 2015

Correspondence to: Elizabeth B. Lewis; e-mail: elewis3@unl.edu
Contract grant Sponsor: National Science Foundation. Contract grant number: 03353469.

INTRODUCTION

Teacher Change Through Professional Development

Since the initial publication of the *National Science Education Standards* (NSES; National Research Council [NRC], 1996) and *Benchmarks for Scientific Literacy* (American Association for the Advancement of Science, 1993) in the United States, teacher educators, professional development (PD) providers, and science teachers have grappled with how to improve student learning and incorporate more inquiry-based instruction in science lessons. The *Next Generation Science Standards (NGSS)* (Achieve, 2013) continues to challenge American teachers with its strong emphasis on not only science concepts but also on scientific practices. As states adopt the NGSS, they will be even more reliant upon classroom teachers who can enact curriculum and instruction that aligns with stated learning objectives and the large-scale assessment that will follow.

With a high value placed upon both scientific knowledge and practices, all students need teachers who can provide meaningful, authentic, and rigorous opportunities to learn science. Additionally, Lee, Quinn, and Valdés (2013) highlighted the pressing need for teachers' science lessons to focus on the language-rich aspects of scientific inquiry and communication for all students that are embedded in scientific practices. They also explicated the need for language support for diverse learners, in particular English language learners (ELL). Thus, it is imperative that science teacher PD programs attend to the wide breadth of knowledge and skills teachers need to enact 21st century science instruction (Bellanca & Brandt, 2010) and meet a modern vision of professional practice (Darling-Hammond & Bransford, 2007).

Because teacher PD is a relatively new idea, only taking root in the 1970s (Lieberman, 1992), it is not so surprising that concurrent production of new science curricula (e.g., Biological Sciences Curriculum Study (BSCS)), without a deep understanding of how to affect teacher change and develop teaching expertise over time, has failed to result in science education reform. Yerrick and Roth (2004) also noted key differences between present and past reform recommendations; in the past, teachers' content knowledge and pedagogy were an isolated concern with little attention to student diversity or learning needs (Lee et al., 2013; Oakes & Guiton, 1995). Over time, PD programs have been more broadly used and diversified, creating myriad options through which teachers improve their science content knowledge, methods for engaging students, familiarity with exciting curricula, knowledge of how to conduct scientific research, and so forth. Despite the popularity of PD, historically the community of teacher educators and in-service PD providers has understood little about exactly how teachers apply what they learn during PD to their classroom practice (Hewson, 2007). However, the existing research about PD programs themselves has led to consensus about six aspects of effective and useful PD programs: (a) a clear focus on classroom practice that involves subject matter and pedagogical knowledge; (b) active and inquiry-based learning; (c) collaborative learning; (d) duration and sustainability; (e) coherence in its goals and design; and (f) school organizational conditions (van Driel, Meirink, Van Veen, & Zwart, 2012). More recently, there has been a greater focus both on conducting research on teacher PD and on improving the rigor of such investigations to address the past lack of understanding.

A major issue with investigating the effects of teacher PD is that while a particular finding might be critical for one program in one context, it may dissimilarly apply to another. Teachers need time to integrate new ideas as they make sense of their own teaching situations at classroom, school, district and state levels. In essence, researchers need to understand teacher learning and the variation in the ways that teachers use what they have learned. Wilson (2013) identified teacher PD as one of the "grand challenges" in

science education research and called for a more complex view of teacher learning, "one in which professional learning is seen as more dynamic and iterative, connecting teachers' experiences in their classrooms with formal opportunities for collective reflection and for acquiring new knowledge that targets genuine problems of practice" (p. 311). In the *Second International Handbook of Science Education* (Fraser & Tobin, 2012), there were three chapters devoted to professional knowledge, science teacher learning, and PD. The authors of one of these chapters, Wallace and Loughran (2012), remarked that connecting teacher learning to school reform is a recent phenomenon, but that "teacher learning is a central tenet for educational reform" (p. 303). To respond to this call for more sophisticated and practical insights into the mechanics of teacher learning and application to the classroom setting, educational researchers will need to carefully align measures and analyses of teacher and student performance to determine how teacher learning translates into teacher effectiveness and to ensure the transferability of findings.

In this study, we investigated community of practice-based science teacher learning as a model for instructional change. We report on the implementation of one such research-based, theory-driven PD program called the *Communication in Science Inquiry Project* (CISIP) designed to help teachers create scientific classroom discourse communities (SCDCs; Baker et al., 2009). These communities use the exploration of the natural world along with oral and written discourse to support learning of core scientific concepts. Through a multimethod, quantitative research design (e.g., surveys and classroom observations of science lessons), we examined the factors that acted as barriers and supports to implementation of SCDCs, which aspects of the PD were adopted more readily than others, and teachers' motivation to change. This information, along with teacher, student, and school demographic information, was then used in the creation of a hierarchical linear model to model change in teachers' implementation of the PD over time.

The teacher PD that we studied leveraged principles of learning in line with traditional learning theory at multiple levels (e.g., students, teachers). Following in the footsteps of Borko and Putnam (1996), we understand that learning to teach draws on cognitive psychology and certain core learning principles: (a) "the central role of knowledge; (b) learning as an active constructive process; (c) knowledge and learning as situated in physical and cultural contexts; and (d) the importance of prior knowledge and beliefs in learning to teach" (p. 673–674). In our investigation of one instance of teacher PD, we use these same core principles to analyze what these particular teachers learned and how they applied what they knew to enact reformed teaching.

LITERATURE REVIEW

Key Aspects of Teacher Professional Development

In the second edition of their book, Loucks-Horsley, Love, Stiles, Mundry, and Hewson (2003) used aspects of effective teacher PD to offer a design framework for PD. These authors synthesized many general but critical aspects of designing effective teacher PD based on their experiences and knowledge of pitfalls to avoid (e.g., insufficient time, recruiting teachers in equitable ways to ensure diversity); thus their book has become part of the essential cannon of the PD provider, especially with a release of the third edition in 2010. In a recent status report on the current state of the field, Wilson (2013) echoed five key aspects of teacher PD that researchers have identified: (a) "focusing on specific content, (b) engaging teachers in active learning . . . , (c) enabling the collective participation of teachers . . . , (d) coherence (aligned with other school policy and practice), and (e) sufficient duration (both in intensity and contact hours)" (p. 310). Van Driel et al.

(2012) also specifically identified school organization conditions as an important, yet understudied, aspect of teacher PD. Indeed, much foresight and planning must be employed to both design research-supported teacher PD and concurrently study the effects of those programs. Van Driel et al. (2012) offer a more current review of research on science teacher PD and have documented the increase in the research literature of studies of science teacher PD. They selectively analyzed 44 studies, ultimately placing them into four categories according to Clarke and Hollingsworth's (2002) model of teacher professional growth: (a) the relationship between external domain and the domain of practice; (b) the relationship between the external domain and the personal domain; (c) relationships among the external domain, domain of practice, and the personal domain; or (d) all relationships, including the domain of consequence (i.e., student outcomes). Across these studies, they identified the fact that researchers frequently did not consider the results of teacher PD in the light of school organizational conditions. Indeed, teacher PD can appear to be more effective by ignoring the practical limitations that teachers may face, which could potentially undermine the positive learning experiences that they have had within a professional learning community. As part of this study we deliberately investigated teachers' perceptions of barriers and supports to implementing PD ideas—in particular, how they viewed their administration, students, students' parents, and colleagues.

A national study by Blank, De las Alas, and Smith (2008) that sampled American mathematics and science teacher PD initiatives from 2004 to 2007 failed to find how observed changes due to PD functioned over time, what changed about teachers' practices, or how to evaluate change over time in a way that aligned theory, methodology, analytic method, and findings. The same report indicated that programs that appeared to change teachers' classroom instruction were over 50 hours in length, but it estimated that only about one third of studies reported measurable effects. Banilower, Heck, and Weiss (2007) conducted a study of National Science Foundation–funded Local Systemic Change projects and found that participation in PD was positively related to attitudes toward, and perceptions of, science instruction, including teaching methods and subject matter knowledge. They also found that teachers were more likely to implement specific instructional materials if they received PD on how to use them. Jeanpierre, Oberhauser, and Freeman (2005) reported that PD for the purpose of shifting secondary science teachers to a more inquiry-based practice ought to include opportunities for practicing science content and process knowledge with teacher accountability. For example, Penuel, Fishman, Yamaguchi, and Gallager (2007) studied teachers engaged in PD with the GLOBE Program, an international earth science education program, and concluded that the success of the GLOBE program included providing teachers with time to generate implementation plans and materials needed for a more inquiry-based approach to learning. Additionally, Penuel et al. (2007) concluded that when providers adapt PD activities to specific groups, they must balance teachers' own contexts, the PD demands, and negotiating PD goals within schools and classrooms.

By acknowledging the complexity of the educational system, this study highlights the need for administrative support for "meaningful experimentation" in school systems as identified by Donovan (2013) to develop a better understanding of how to reform education. Like Hewson, O'Donnell (2008) reminded us there is insufficient research to guide researchers on "how fidelity of implementation to core curriculum interventions can be measured and related to outcomes, particularly within efficacy and effectiveness studies, where the requirements for fidelity measures differ" (p. 33). When administrative policies and research goals are at odds, or access to schools is prevented, we are unable to investigate how teaching innovations work in real classrooms across multiple contexts with diverse students.

As a closing point, the assumption is that student performance is generally correlated with teacher effectiveness and increased teacher effectiveness with more PD. However, because every instance of PD is idiosyncratic, global claims about all PD are difficult to make. Fidelity to PD and its similarities in implementation to other programs is critical to making larger claims about overall traits of PD that are correlated with student learning gains. Nevertheless, we need well-vetted innovations, and to have such innovations, we must have a clearer understanding of how PD is incrementally adopted and implemented or rejected.

Conceptualizing Teacher Change: Learning Theory and Communities of Practice

Kunzman (2003) identified five themes within experienced teachers' learning: (a) a greater awareness of struggling students, (b) more complex understanding of curriculum planning, (c) the importance of collegiality and collaboration, (d) value of feedback and structured reflection, and (e) development of a theoretical framework to inform and guide practice. Such aspects of teachers' learning are often identified as cornerstones to good teaching (Darling-Hammond & Bransford, 2007). The importance of collaboration and collegiality to support community-based situated learning and practice supports sociocultural theories of learning (Lave & Wenger, 1991; Vygotsky, 1986). There are many aspects of learning (e.g., cognitive, affective, motivation) that can be used to understand teacher change. In the past, researchers like Borko and Putnam (1996) framed their synthesis of research findings around teachers' beliefs, subject matter knowledge, and general pedagogical knowledge. The essential quality of a classroom is in the interactions among these categories and other factors; therefore, limiting findings to isolated categories is inevitably an oversimplification. To avoid unwarranted findings, the use of core learning principles must point directly to the particular mechanisms by which learning occurs. In our investigation, we used cognitive learning principles to analyze science teachers' learning by focusing on how they applied new knowledge to enact reformed teaching, thus examining changing instruction in its complexity. Specifically, we employed the following conceptual framework to design a study to better understand how teachers learned how to build scientific classroom discourse communities (SCDC) through PD (Figure 1).

In this view, classrooms are ecosystems, subcultures, communities of practice, places of social reproduction, and microcosms of the communities within which they are situated. Science teachers must navigate their own professional goals, the daily demands of students, parents, colleagues, administrators, and workplace cultures. In the same way, students navigate their own intersecting, complex milieus. There has been a convergence in the research literature on teacher and student learning highlighting their similarities (Loughran, 2007). Teachers may learn new ideas through PD, but may implement them selectively because of their erroneous beliefs about students and how they learn (i.e., intelligence is a fixed quantity, not changeable [Dweck, 2000] and thus only highly motivated honors students can be challenged with inquiry-based science instruction, rather than all students). Similarly, students may learn new scientific ideas and adopt, or not adopt, them based upon their personal beliefs. We used psychological theories of individual cognition to frame both the content of the PD and our study of teachers' learning (Table 1). The three core learning principles are (a) engaging prior understandings, (b) the essential role of factual knowledge and conceptual frameworks in understanding, and (c) the importance of self-monitoring (e.g., metacognition; NRC, 2000, 2005). Our application and research design using these principles will be explained in greater detail in a later section.

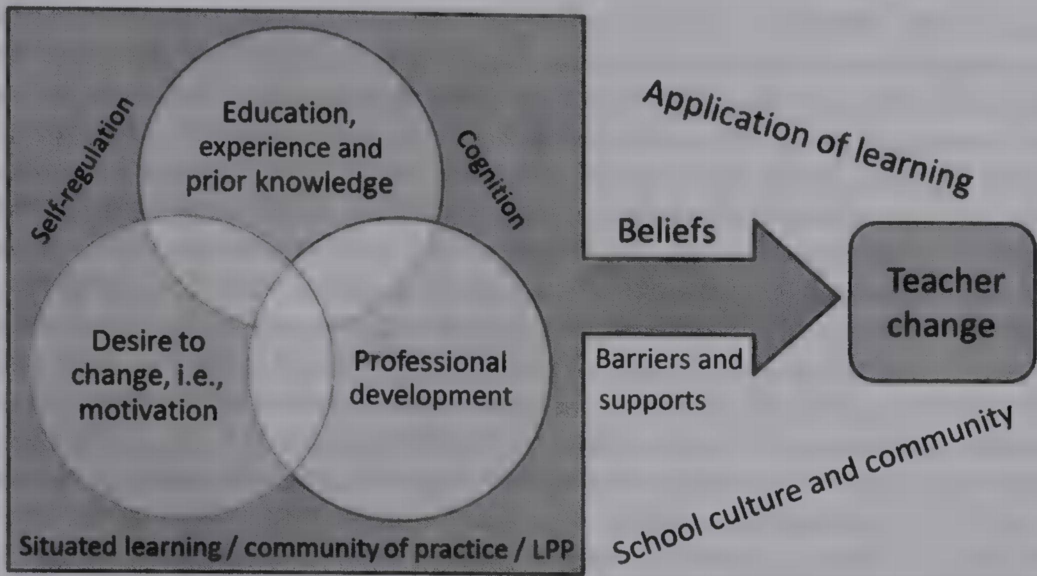


Figure 1. Model conceptual framework of teacher learning and change through cognition, self-regulation, that corresponds with cognitive learning principles and situated learning with respect to individual values and institutional contexts.

TABLE 1
Matrix of Learning Principles, Teacher's Learning Through PD, and Instrumentation

Learning Principles (NRC, 2005) Student Learning	Teachers' Learning Through PD	Instruments Used To Generate Data
LP 1: Engaging prior understandings	Prior knowledge of instructional strategies, beliefs, science content knowledge, credentials, pedagogical knowledge	• Teacher education and demographic survey • *CISIP Teacher Self-Reflection Survey*
LP 2: The essential role of factual knowledge and conceptual frameworks in understanding (and assessment of this knowledge)	Facts = individual and observable instructional strategies Conceptual framework = CISIP model of a scientific class discourse community (SCDC) • inquiry • oral discourse • written discourse • academic language development • learning principles	• Descriptive statistics individual instructional strategy use within framework of SCDC • Observations of teaching (DiISC instrument) = authentic/performance assessment of learning • Change in enacted practice (HLM)
LP 3: The importance of self-monitoring	Teachers reflection and identification of what supports and prevents (barriers) their implementation of a model of a SCDC	• *Barriers and Supports Survey*

More broadly, Vygotsky's (1986) social development theory of cognition emphasizes the pivotal role of culture, language, and social factors. The concept of a zone of proximal development (ZPD) explains how more capable learners can provide the necessary scaffolding for new or struggling learners. So, in addition to learning theory that focuses on individual cognition, we used the concept of ZPD by having master teachers mentor new teachers within a community of practice (using, as we call it, reciprocal teaching methods). This idea is well outlined by Lave and Wegner (1991) who studied apprenticeship as a mode of learning, developing ideas of situated learning and communities of practice. In particular, their concept of legitimate peripheral participation (LPP) required mentoring of novice members. Student teaching is analogous to apprenticeship in the current model of preservice teacher education, and participating in teacher professional learning communities is the emergent model for in-service teacher PD. Parallels between ZPD and LPP reinforce each theory of learning in social contexts, and many educational researchers have used these theories, thus adopting a situated learning perspective (e.g., Putnam & Borko, 2000). In light of powerful social forces, Lave and Wenger (1991) developed an analytic perspective for educational researchers; situated learning bridges both individual cognitive processes and group social practices, allowing researchers to capture the complexity of the phenomenon of teacher change. From these ideas emerges the concept of a scientific classroom discourse community (Hand et al, 2003; Yerrick & Roth, 2004) to more authentically match the practices of scientists and provide more engaging opportunities to learn science. In this study of teachers' learning and changing practices, we applied Lave and Wenger's analytic viewpoint as others had successfully done (e.g., Franke, Kazemi, Carpenter, Battey, & Deneroff, 2002) to specifically study teachers' participation in PD activities focused on learning how to build their own scientific classroom discourse communities.

Language, Learning Science, and Scientific Classroom Discourse Communities

Science education reform documents (Achieve, 2013; NSES, NRC, 1996) have encouraged science teachers to use authentic learning experiences that reflect the ways in which scientists communicate their own work. Scientists work in teams of researchers, peer-review each other's work, and communicate their findings through a variety of oral and written modes. Thus, to better reflect the practice of doing science, science teachers need to be able to bridge these uses of academic language and practices of scientists with students' everyday language and conceptions of the world around them.

Lemke's (1990) identification of classroom triadic dialogue (initiate–respond–evaluate, otherwise known as "IRE") as a means for knowledge transmission and discourse structure is the antithesis of science education reform. However, Lemke found that it is a favored staple of whole-group discussion pedagogy in science classes. The use of scientific inquiry as a teaching paradigm provides students with more opportunities, not only to engage with scientific questions, make observations, and make meaning from their own experiences, but also to talk with each other and not just their teacher. Gee (2005) stated that students need these peer-to-peer learning experiences to create meaningful discourse and develop conceptual understandings. This follows in the Vygotskian (1986) and Dewian (1938) tradition of social and experiential learning and language. Numerous authors have written about the sociocultural, sociolinguistic, and philosophical elements of scientific classroom discourse communities and the importance of language in learning science (Yerrick & Roth, 2004). For example, in *The New Science Literacy* (Their and Daviss, 2002) and *Crossing Borders in Literacy and Science Instruction* (Saul, 2004) the authors illustrate a

combination of science, language, and learning that are now on the leading edge of science education reform. The CISIP PD program relied heavily upon the use of language and learning theories in developing its model of a scientific classroom discourse community; with this model in hand, one of the main goals of the teacher PD was for teachers to learn how to address the needs of their diverse learners and underrepresented students in science.

As Borko (2004) reported in an analysis of PD research, "we have evidence that PD can lead to improvements in instructional practices and student learning" (p. 3). This conclusion is encouraging and by researching the critical elements of PD that can foster educational reform we can be more effective in providing teachers with opportunities to adopt new practices. In this study, our main objective was to understand how teachers applied a specific PD model as they designed new curriculum and implemented a wider range of instructional practices, focusing specifically on how they constructed scientific classroom discourse communities. We also investigated impediments and supports to teachers' transformed practices. Within classroom discourse communities, we examined the complex relationships embedded within teaching as a social act and as more than a simple set of behaviors (Erickson, 1986; Lave & Wenger, 1991). We explicitly highlighted and used scientific classroom discourse communities in the PD to model how science and English language arts/ELL teachers could approach teaching and learning with their own students.

RATIONALE AND RESEARCH QUESTIONS

Our study investigated the issue of science teacher reform through changes in instructional practices. In this case, the PD program focused on learning about a set of instructional strategies from which teachers could chose to design their own scientific classroom discourse communities. This PD design hinged upon salient research findings and the practical needs of science teachers, following a pragmatic perspective which has been espoused and synthesized by Wallace and Loughran (2012). They comment that a pragmatic perspective "would suggest that teachers need the opportunity to engage in authentic activities, participate in rigorous and critical debate within discourse communities, and develop facility with the various tools used in that community" (p. 302). The PD program design and setting in this study encompassed aspects of individual cognition, social interaction, and the learning environment. These variables are dynamic, which complicate studying how teachers learn from specific PD programs, reflect on their teaching practices, and selectively implement what they have learned in their classrooms. Thus, in many ways all research about specific, unique PD experiences will be highly contextualized at two levels: the general level of the PD program design and the more specific level of what will be incorporated into the classroom by different participants.

Throughout our study, we found that fidelity of implementation is a double-edged sword; sometimes it is difficult to balance respect for teachers as experts in their classrooms with outcome-driven PD agendas, but we assumed that effective PD would improve teachers' knowledge to the extent that it could be observed as a change in their classroom instruction. Table 1 aligns learning principles, teacher learning through PD, and the instruments we used to generate data. We asked the following research questions as part of our overall inquiry into teacher implementation of PD:

1. Which of the instructional strategies from the CISIP did teachers adopt more easily than others to create their own scientific classroom discourse communities?
2. What, if any, student or teacher variables significantly predicted teachers' implementation of the CISIP model or their initial levels of PD-associated behaviors?

3. To what degree were science teachers motivated to change their instruction to be more aligned with the CISIP model?
4. What were teachers' views of barriers and supports to implementing new ways of teaching science?

PROFESSIONAL DEVELOPMENT RESEARCH PARTICIPANTS AND CONTEXT

Teachers were recruited into the PD program in school-based teams with administrator support. Districts were approached initially to determine their interest before recruiting teachers; in fact the administrators were also provided with a 1-day PD session to learn more about the CISIP PD activities so as to better understand the kinds of changes teachers might be making in their classrooms. The teachers were provided with an honorarium to participate during the summer sessions and follow-up Saturday workshops throughout the school year. The majority of the teachers who started the CISIP program stayed with it from beginning to end, but there was approximately a 15–20% attrition rate. During the first year, middle and high school teachers participated in one of two 3-week CISIP summer institutes, followed by 4 day-long workshops to reinforce and elaborate upon the summer PD (Figure 2). The teachers had an opportunity to attend a total of 96 hours of PD programing in the first year. Some teachers had also previously participated in the 2-year development phase, and potentially had accrued an additional 200 hours. During the second year of the study, only high school science and English language arts/ELL teachers from two school districts were observed. Teachers who had participated in the first year acted as mentors and recruited new teachers. These new teams participated in a 4-day introduction to CISIP over the summer and six workshop days throughout the academic year, for an additional 60 contact hours.

The research team was separate from the PD program team, but interfaced regularly with the PD providers to provide feedback from not only the classroom observations between

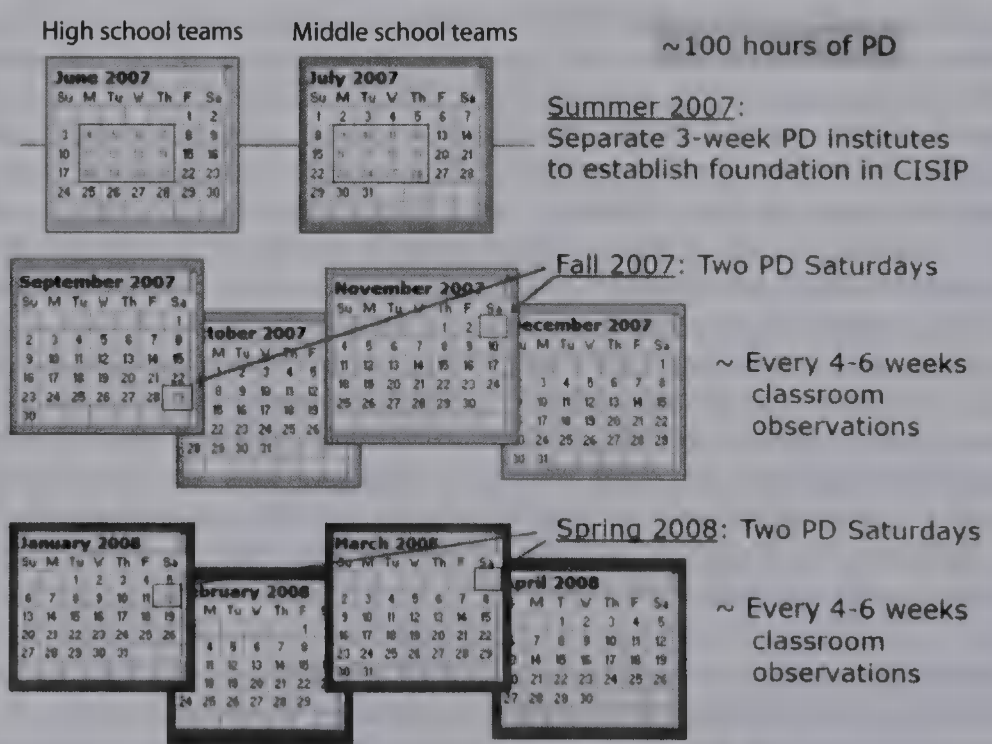


Figure 2. CISIP professional development schedule.

PD sessions but also from the PD workshops. We acted as unobtrusive observers in the classroom when we made observations of the teachers, and we did not provide coaching as there were teacher-leaders already in place to provide support. Our focus was mainly on the teaching behaviors of the teachers themselves and what, if any, aspects of the PD they were trying to use. There was also an external evaluator on the grant who worked independently of the research team but occasionally interacted with the researchers to compare fieldnotes and provide annual reports and feedback to the lead PD providers, principal investigators, and the grant's advisory board.

The *Communication and Scientific Inquiry Project* Community of Practice

While we, as part of the research team, were interested to see how much of the PD from a specific program was used by teachers, the CISIP program itself rejected the notion of scripted science lessons. While this kept PD context specific to individual teachers' practices, our findings helped guide the development of tools for science education reform. The goal was to teach secondary science teachers how to build SCDCs from a wide range of aligned instructional strategies. Teachers were encouraged to develop their capacity through the development of an "instructional palette," used in turn to design lessons to meet diverse students' learning needs. Teachers had the opportunity to (a) learn more about effective teaching methods, (b) practice designing and teaching science lessons, and (c) confront negative beliefs about teaching science to all students.

At each of the PD sessions, teachers were provided with exemplar activities using specific instructional strategies to model particular aspects of a scientific classroom discourse community. They participated in these activities themselves and then were provided time in groups to brainstorm ways that they could use those same instructional strategies in the context of their own curriculum and students. For example, in the *Mystery Boxes and the Writing of a Scientific Explanation Activity* teachers were provided sealed wooden boxes with objects inside and were asked to generate observations and construct claims using evidence and reasoning. This activity modeled for the teachers the writing process of a scientific explanation with an emphasis on clear performance expectations for writing and the writing of an explanation with claims, evidence, and reasoning. They were also provided feedback on written scientific arguments and revising arguments based upon their teams' writing to model another critical aspect of student learning. This example aligned most strongly with the SCDC aspect of written discourse. Other examples of the four other aspects are presented in Table 2.

The CISIP community of practice included beginning and veteran teachers, in-service teachers, secondary and postsecondary science teachers, and English language arts and ELL faculty. English language arts and ELL teachers were included as part of the school-based teams because of their expertise in oral and written discourse, and it was conceived that they could assist their science colleagues in these areas. The range of teacher knowledge made all teachers simultaneously experts and novices in an interdisciplinary teaching dialogue that drew upon available expertise. The CISIP participants were part of a teacher learning community as defined by Cochran-Smith and Lytle (2003) as "social groupings of new and/or experienced educators who come together over time for the purpose of gaining new information, reconsidering previous knowledge and beliefs, and building on their own and others' ideas and experiences... intended to improve practice and enhance students' learning" (p. 2462). All teachers in the PD had something to learn from each other because the CISIP model was built upon and integrated critical aspects of multiple disciplines to benefit both nascent and master teachers. Thus, situated cognition

TABLE 2
Selected CISIP Professional Development Activities for Teachers to Learn to Build Scientific Classroom Discourse Communities

SCDC Core Elements	Activity Example
Scientific inquiry	• *BioLab 1: Human Characteristics*: Inquiry investigation about human characteristics with embedded support for academic language development with modeled strategies to use in the classroom. • *BioLab 2: Gummy Bear Genetics*: Experience and use of academic language development strategies embedded within an CISIP inquiry activity about genetics. • *BioLab 3: DNA Extraction*: Integration of CISIP components within DNA laboratory.
Oral discourse	• *Nature of Science (NOS) Communication Card Activity*: Definition of NOS and the types of communication that are integral to doing science. Discussion about how scientific writing and talking reflects NOS
Written discourse	• *Mystery Boxes and the Writing of a Scientific Explanation*: Begin writing process of a scientific explanation with an emphasis on clear performance expectations for writing and the writing of an explanation with claims, evidence, and reasoning. Provide feedback on written scientific arguments and revise arguments based upon writing.
Academic language development	• *Opening Doors*: Experience and identification of scaffolding strategies and techniques for teaching academic skills to English language learners (ELL). • *BICS/CALP*: Explanation of the significance of Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP) in language acquisition.
Learning principles	• *Fish is Fish*: Introduction to learning principles and the sociocultural influences on ELL as they relate to "Fish is Fish" story. • *Graphing Motion with Motion Detectors*: Situating of metacognition within an inquiry activity. Development of concepts of graphing of back and forth motion with attention to metacognition.

and LPP were foundational and the learning community encouraged sharing of subject matter knowledge and instructional approaches from each discipline.

Based upon critical research findings, the CISIP model included five essential curricular aspects to design effective science instruction: (a) scientific inquiry, (b) oral discourse, (c) written discourse, (d) academic language development, and (e) learning principles (e.g., accessing student's prior knowledge (NRC, 2000, 2005)). As a learning platform, scientific inquiry that relied upon a constructivist learning approach provided teachers with opportunities to engage with scientific questions, make observations, and interpret data to generate their own conclusions in the same ways as their students. While the instructional

strategies promoted in CISIP were carefully selected from relevant research literature, the types of lessons teachers designed for students ultimately determined what, if any, benefits students gained as a result of their teachers' PD. Teachers were regularly provided time to cogenerate lessons with colleagues. Over time, the PD providers collected and shared teacher-generated examples of transformed lessons using the CISIP model.

In summary, CISIP provided school-based teams of teachers with year-round PD that regularly focused on (a) ELLs' needs and the challenges of academic language acquisition for mainstream students, (b) opportunities for teachers to redesign lessons using SCDC instructional strategies, (c) activities for teachers to exchange ideas, (d) opportunities for teachers to reflect upon their own learning during activities, and (e) regular and explicit instructional examples and connections to the SCDC model. The CISIP PD model also included rigorous use of student science notebooks with embedded academic language learning support. The PD program carefully wove the aforementioned five core elements throughout the activities for the teachers (see Table 2 for selected examples). Over time the PD providers collected and showed teacher-generated examples of lessons that had been transformed using the SCDC model. Teachers were also provided time during the PD to develop their own lessons for their own students.

METHODOLOGY

Participants

Of the teachers participating in the CISIP PD, there were a total of 16 high school and 13 middle school teachers, mostly female (69%), with an average of 11.3 years (SD = 8.9 years) of teaching experience, who consented to allow classroom observations. Their demographic information is given in Table 3. Included in our entry survey of teaching demographics, we also asked teachers to provide us with some indicators of their prior knowledge, e.g., how to teach ELLs, science methods coursework, course(s) on the history and nature of science (NOS; e.g., 48% of teachers without), thus providing some indicators of what teachers might know about the CISIP core ideas prior to starting the program.

Data Collection and Researcher Stance

There were three levels of our investigation: (a) *Level 1*: surveys of 11 middle and 14 high school science teachers who participated in first year of CISIP, (b) *Level 2*: 15 middle and high school science teachers who consented to regular classroom observations, and (c) *Level 3*: the classroom instruction and perceptions of PD of two high school biology teachers. The data collection timeline was as follows: (a) Upon their entry into the PD program, teachers were asked to fill out a demographic questionnaire and a beliefs survey; (b) as they engaged with the PD, we scheduled four to six observations throughout the school year; and (c) at the end of the PD program, we had them take the belief survey again (postprogram) and complete a survey of what they viewed as supports or barriers to implementing the PD in their own classroom.

When we conducted observations, we generated fieldnotes that described the focus and science content of the lesson that were covered, the classroom activities that occurred, the kinds of instructional strategies that were being used by the teachers, and the kinds of discourse that were occurring (e.g., small group, whole group). We did not transcribe the lessons, as we did not intend to engage in linguistic discourse analysis, but rather classified the types of discourse instructional strategies that occurred (e.g., peer to peer). We also collected copies of any handouts that the teachers provided their students. These fieldnotes

TABLE 3
Teacher Education and Demographic Information

Teacher Demographic Information	
Middle school	13
High school	16
Female	20 (69%)
Male	9 (31%)
Average years teaching	11.3 ($SD = 8.9$)
Average number of degrees	1.76 ($SD = 0.64$)
Bachelor's degree	9 (31%)
Post-Baccalaureate course work	4 (13.8%)
Master's degree	15 (51.7%)
Medical Doctorate degree	1 (3.4%)
Certification	
No teacher preparation	1 (3.4%)
Undergraduate teacher certification program	12 (41.4%)
Postbaccalaureate teacher certification program	16 (55.2%)
In-field	27 (93%)
Out-of-field (elementary)	2 (6.9%)
PD-relevant coursework	
Mean number of science courses	17.3 ($SD = 10.49$)
Mean number of science methods courses	1.7 ($SD = 2.00$)
Teachers without a class in history and philosophy of science	14 (48%)
Teachers without an English content course	7 (24%)
Teachers who had—one to two English classes	8 (27.5%)
Teachers without an English or language arts teaching methods class	17 (58.6%)
Teachers without an ESL class	5 (17.2%)

allowed us to use the *Discourse in Inquiry Science Classrooms* (DiISC) instrument to determine the degree of alignment with the CISIP model of a scientific classroom discourse community.

Over 2 years, the research team conducted 297 observations of teachers' science lessons; the distribution and participants in these observations are as follows: In the fall of 2007, the lead author observed 14 Level 2 teachers one to four times each for a total of 31 observations; she conducted most of these observations with another researcher and engaged in interrater consensus discussions after each observation. Other members of the research team also made other observations in pairs. In the spring 2008, the lead author, who was also primarily responsible for the training of other observers, observed six teachers one to ten times each for a total of 24 solo observations; other members of the research team also made solo observations. Thus, during the 2007–2008 academic year, 106 classroom observations of CISIP science teachers (Level 2) were conducted. We then used the observation scores to build an exploratory, 1-year longitudinal model using hierarchical linear modeling (HLM) to determine what, if any, significant relationship existed between various teacher attributes and teachers' fidelity to the CISIP model (Lewis, 2009).

Because the results of the 1-year HLM were tentative, we generated another year of observation data to build a better-powered model; these results are presented below. Over

the course of the 2008–2009 academic year, we made an additional 163 observations (first in pairs, and once reliability was reconfirmed, as independent observers) of 10 original participants and 16 of their newly recruited teaching colleagues for a total of 30 teachers (16 science and 14 English language arts/ELL teachers). Seven of the 10 original teacher participants had previously participated in the PD but not in the research study. Additionally, we made 28 observations of 13 comparison (i.e., non-CISIP) science teachers. The lead author also constructed case studies of two high school biology teachers (Level 3) that are presented elsewhere due to space constraints (Lewis, 2011).

The research team was also part of the instrument development team that engaged in extensive field-testing and constant comparison with the CISIP program sessions. We developed this instrument because there were few available classroom observation instruments at the time and none that were aligned with the content of the PD. For over a year, the research team conducted observations in pairs and generated consensus scores and refined the items to be unidimensional. After determining that interrater reliability had been achieved, the observers conducted observations independently.

Instruments

Each of the 323 teachers' lessons was scored with the DiISC instrument. The DiISC was developed over 3 years and was aligned with the SCDC model; its development is chronicled in greater detail elsewhere (Ozdemir, Lewis, & Baker, 2007). Of note is the fact that we have not, as of yet, established a holistic validity and reliability argument for using this instrument. Initially, the items were developed in reference to previous research on the role of writing, oral discourse, scientific inquiry (NRC, 1996), learning principles in science teaching and learning (NRC, 2000, 2005), and academic language development strategies. A manual for use with the DiISC was developed and outlines the theoretical underpinnings of the development of the instrument as well as the psychometric properties (Baker et al., 2008). The five scales on the DiISC match the five aspects of CISIP. We used the 36-item DiISC as proxy for teacher fidelity to the CISIP model, to better understand which instructional strategies were used more often than others, and model teacher change over time. Each item used a 0–3 point scale with a unique rubric. To reiterate, based on an insufficiently developed validity argument (due to time and sample size limitations), we used proportional scores (total teacher score/total possible score) within the five scales, rather than a more complex composite score (e.g., principle components analysis). In fact, attempts to simplify the response patterns (using principle components analysis) or examine underlying factor structure (using exploratory factor analysis) yielded results that were uninterpretable. Other work is being done to improve this measurement device and generate proper, holistic validity and reliability arguments, but until that work is completed, we are unable to make the case that results similar to those we found would be possible without using the DiISC in the same way as it was used in this study.

Two exploratory surveys were implemented, the *CISIP Teacher Self-Reflection Survey* (comparing teachers' current and desired use of CISIP), and a survey of barriers and supports to implementing PD. We also used an educational history and teacher demographics questionnaire to complement the classroom observations as a means to investigate teachers' motivation to change, their learning from the PD and how they used instructional strategies to build their own SCDC, and what factors appeared to support or confound teachers' efforts to change their practice. The *CISIP Teacher Self-Reflection Survey* was written in an effort to determine teachers' desire to change their instructional practices and included 19 Likert-type items aligned with the five CISIP aspects and one item on lecturing, which was a teaching method that the PD sought to decrease in its frequency.

The 30 science teachers in the study took this survey before they started the summer institute. Teachers rated the frequency of occurrences of different teaching methods within their classrooms from two perspectives: "the way it is," and "the way I'd like it to be." The survey had a repeated-measures design, and data were nonparametric; we conducted sign tests to identify significant differences between the medians of the sampled teachers' current and desired teaching practices on any of the 20 items. We used Kendall's *tau-b* to measure the degree of correspondence between each pair of teachers' ratings and to assess the significance of this correspondence in an effort to determine whether there was a statistical relationship between each pair of variables for teachers' current and desired frequency of a specific instructional strategy.

Science teachers responded to a second survey designed to assess their perceptions of various categories of barriers and supports to PD implementation. We designed the survey based on teacher comments as well as a systematic list of variables that could potentially affect teachers' views toward implementation. The survey categories were (a) administrative actions, (b) collaborative teacher relationships, (c) curriculum, (d) instruction, (e) parents, and (f) students. This 46-item survey used a five-point Likert scale, rating major to minor supports for implementing PD. The items were tallied by subgroups, and the means were calculated. We set ranges between 1.0 and 5.0 to classify the groups' mean response to each item as a barrier (1.00–1.50 = major barrier; 1.51–2.49 = minor barrier) or a support (4.50–5.00 = major support; 3.50–4.49 = minor support) to obtain a rough approximation of teachers' perceptions as a group within each category.

These surveys were meant only to provide exploratory, descriptive results. There has been no development of a validity and reliability argument associated with these surveys, as they are not related to inferences we make here. The reliability and validity arguments of the DiISC were not adequately developed for our purposes to generalize findings, and the surveys were meant only to be descriptive. Our goal was to build credible findings not generalizable inferences, and thus we: (a) provide descriptive information for other researchers, (b) establish research questions that can be investigated with greater rigor in other studies, and (c) characterize the specific results of this study.

Modeling Teacher Change Over Time

We used HLM to explore relationships between PD, teachers' practice, and systemic variables (Raudenbush & Bryk, 2002; Shadish, Cook, & Campbell, 2002). We chose to use a Hierarchical Linear Model (HLM) for several reasons. Primarily, because we were unable to meet the assumptions of analysis of variance (ANOVA) or other, related general linear model techniques. Conversely, we did not have the sample size to conduct a multilevel Structural Equation Model (SEM) without making the assumptions that would transform the SEM into an HLM. Because HLM is technically a type of SEM, and the assumptions of our analysis reduced the SEM to an HLM, we will refer to our modeling process as using only HLM. We also used HLM because our sample had missing data over time (unequal sample sizes at each time point).

We used several variables to account for initial differences between student groups and treatment over time. We chose these variables by creating an exhaustive list based on available information. As such, the analysis was exploratory. With this technique, individuals can be clustered within time points, so that the number of individuals at any time point could change (Raudenbush & Bryk, 2002); this was needed as teachers joined and exited the study at different times with more or less PD. Our sample size also required that we use a linear rather than nonlinear model. In the construction of the model, we used available teacher demographic information on professional experiences (e.g. length

TABLE 4
Summary of 1 Year of DiISC Observations ($n = 106$) of All Science Teachers ($n = 16$)

Scale	Number of Items	Maximum Score	Median	*M*	*SD*	z-Scaled Mean	z-Scaled *SD*
Scientific inquiry	6	18	3.0	3.36	3.24	0.19	0.18
Oral discourse	5	15	5.0	5.37	3.11	0.36	0.21
Written discourse	6	18	4.0	4.50	2.44	0.25	0.14
Academic language development (ALD)	8	24	7.0	7.51	3.22	0.31	0.13
Learning principles	11	33	7.0	7.72	4.10	0.23	0.12

of time teaching). We selected eight additional variables for their potential correlation with teachers' implementation of PD (Cuban, 1992), including, but not limited to school district size, per pupil spending on classrooms, total spending costs, socioeconomic variables, and average teacher salaries for each teachers' district (data source: [State blinded for anonymity] Department of Education, 2008). We used the DiISC scores as our outcome measure.

For the longitudinal model, our sample size allowed a two-level model (we attempted a third level, but the model was underpowered). The first level included the total raw observation scores on the DiISC for all five areas. The second level included a dummy code for group participation (PD or non-PD comparison group) with demographic information. Ultimately, only the two models described below allowed us to make inferences with statistical evidence. With a small, contextualized sample size, our investigation was exploratory and limited our capacity to generalize to other groups of teachers in the larger population or definitively decide between the two final models.

RESULTS

Below, we present the results of which CISIP instructional strategies teachers used over the course of the first year of PD, as well as the results of a 2-year HLM to show how teachers' instruction changed. Finally, to explore the possible reasons behind these changes we conclude with summaries of results from the *CISIP Teacher Self-Reflection Survey* and *Barriers and Supports Survey*.

Research Question #1: Teachers' Adoption of SCDC Instructional Strategies

During the first year of PD, we found that teachers' use of the CISIP scientific classroom discourse community model varied in implementation (see Table 4). On each scale, the science teachers, based on a comparison of their z-scaled means, scored from highest to lowest in their use of groups of strategies: (a) oral discourse, (b) academic language development, (c) written discourse, (d) learning principles, and (e) scientific inquiry. The means were used to rank order all teachers' ($n = 16$) use of the CISIP instructional

TABLE 5
Rank Order by Mean of Most to Least Used CISIP Instructional for All Science Teachers

Scale	Item	Description	Mean	*SD*
ALD	20	Clear instruction	2.11	0.83
Writing	18	Use of notebooks	1.50	0.93
ALD	19	Vocabulary acquisition	1.43	0.78
Oral	11	Model science discourse vocabulary	1.38	0.80
ALD	21	Visual aids gestures	1.38	0.79
Oral	9	Small group discussion	1.35	0.95
LP	42	Feedback	1.32	0.79
LP	38	Community norms	1.24	0.83
LP	39	Teacher expectations	1.17	0.72
LP	32	Review concepts	1.11	0.87
Oral	10	Bridge everyday with academic	1.06	0.91
Writing	14	Prewriting	1.05	0.87
Sci. Inq.	1	Inquiry environment	1.04	0.94
Oral	8	Whole-group divergent questions	1.04	0.82
Writing	16	Practice scientific writing	1.03	0.79
LP	31	Facts and conceptual framework (NRC, 2005)	1.03	0.79
ALD	25	Organize groups structure roles	0.85	0.85
LP	34	Metacognition (NRC, 2005)	0.76	0.86
Sci. Inq.	4	Observe/data collection	0.73	0.97
ALD	22	Bridge language and culture with science	0.63	0.77
Sci. Inq.	5	Claims-evidence	0.59	0.92
Oral	12	NOS discussion	0.55	0.87
ALD	24	Direct instruction learning strategies	0.55	0.74
Sci. Inq.	2	Students ask questions for investigation	0.46	0.78
Writing	13	Formal scientific writing	0.46	0.78
Writing	17	Writing instruction	0.41	0.67
LP	28	Assessing prior knowledge (NRC, 2005)	0.32	0.68
Sci. Inq.	3	Design exploration	0.28	0.64
ALD	23	Differential instruction language	0.28	0.53
ALD	26	Available supplementary resources	0.27	0.67
LP	35	Self-monitoring	0.27	0.61
Sci. Inq.	6	Data interpretation / sources of error	0.25	0.69
LP	37	Executive control	0.25	0.66
LP	36	Self-awareness	0.18	0.45
LP	29	Modifies instruction	0.07	0.29
Writing	15	Rubrics for revision of writing	0.06	0.23

Abbreviations: ALD, Academic language development; LP, learning principals; NOS, nature of science; Sci. Inq., scientific inquiry.

strategies to see which elements of CISIP were used most and least (Table 5). Generally, the teachers' frequency of use of these strategies within lessons fit into three categories: (a) *most-observed* (*often-* and *sometimes-used*) strategies that required teachers to change their own communication, classroom management, and direct instructional behaviors; (b) *occasionally observed* strategies that provided opportunities for greater oral and written discourse to facilitate students' meaning making of science; and (c) *least observed* strategies that encouraged students' executive control of their own learning and teachers' use of formative assessment to be more responsive to students' diverse learning needs (Table 6).

TABLE 5
Frequency of Use of Instructional Strategies through First Year of PD

Scale	Often Used ($M = 1.51 +$)	Sometimes ($M = 1.01–1.50$)	Occasionally ($M = 0.51–1.00$)	Rarely Used ($M < 0.50$)
Scientific inquiry (SI)		SI 1 inquiry environment	SI 4 observe/data collection SI 5 claims-evidence	SI 2 students ask questions for investigation SI 3 design exploration SI 6 data interpretation/ sources of error
Oral discourse (OD)		OD 8 whole group divergent questions OD 9 small group discussion OD 10 bridge everyday with academic OD 11 model science discourse vocabulary	OD 12 Nature of science discussions	
Written discourse (WD)		WD 14 prewriting WD 16 practice scientific writing WD 18 use of notebooks		WD 13 formal scientific writing WD 15 rubrics for revision of writing
Academic language development (ALD)	ALD 20 clear instruction	ALD 19 vocabulary acquisition ALD 21 visual aids gestures	ALD 22 bridge language and culture with science ALD 24 direct instruction learning strategies ALD 25 organize groups' structure roles	ALD 23 differential instruction language ALD 26 available supplementary resources
Learning principles (LP)		LP 42 feedback LP 38 community norms LP 39 teacher expectations LP 32 review concepts LP 31 facts and conceptual framework	LP 34 metacognition	LP 28 assessing prior knowledge LP 35 self-monitoring LP 37 executive control LP 36 self-awareness LP 29 modifies instruction

These groups are described in more detail as follows.

Most-Observed PD Strategies

The subcategory of *often-used* instructional strategies was solely occupied by an underlying stance of clear instruction by modeling expectations in the set of academic language development strategies (ALD, #20), which topped the list by far with a mean of 2.11 ($SD = 0.83$); nearly in a category of its own. A "2" score on this item indicated that the "teacher provided clear objectives and directions" to the students. To score a "3," teachers would have to have been observed monitoring students for their understanding of objectives and directions. In practice, we observed that teachers provided students with clear objectives and directions in their lessons and that some teachers used monitoring more consistently than others.

Included in the subcategory of *sometimes-used* PD strategies (item means ranged from 1.03 to 1.50) were 15 items from all five scales. These strategies included two ALD (vocabulary acquisition and the use of visual aids and gestures to support scientific language), one inquiry (establishing an inquiry environment), four oral discourse (modeling scientific discourse and vocabulary, small group discussion, bridging everyday with academic language, and asking more divergent questions of the whole group), three written discourse (use of science notebooks, prewriting, and practicing scientific writing), and five learning principle items. For example, as they engaged in the CISIP program teachers began to use science notebooks more often, which provided students with a place to record their ideas and engage in prewriting. Teachers also employed small-group discussion more frequently and used more divergent questions when they conducted whole-group discussions. Such instructional moves were also a step toward using more inquiry-based instruction.

Occasionally Used Strategies

We less frequently observed seven other strategies that were at the crossroads of scientific inquiry, discourse and NOS (item means ranged from 0.55 to 0.85). We occasionally observed students collecting data and making claims supported with evidence, discussing NOS, and using metacognition to reflect upon their learning. Teachers occasionally used some critical academic language development strategies, such as assigning students roles within small groups, providing direct instruction about learning strategies, and bridging students' language and culture with the academic register of science. Because these were science teachers with little formal education in the use of language arts, they may have lacked the awareness and confidence to employ such strategies on a more regular basis without further mentoring.

Least Used PD Strategies

Despite regular PD sessions, teachers still struggled with using strategies that placed more choice (e.g., executive control) and self-regulation (e.g., self-monitoring and self-awareness of learning) in students' hands (item means ranged from 0.06 to 0.46). For instance, a student-designed open inquiry-based investigation in which students generated their own research questions and procedures was a rare occurrence in these teachers' curriculum. Students were rarely encouraged to find sources of error in their investigations and engage in formal scientific writing with rubrics for revision of their own writing. Finally, teachers were rarely observed to use formative assessment to revise their instruction.

However, while teachers used more guided than open inquiry instructional methods in their classrooms, they began to noticeably change their instruction. While this overall, 1-year use of CISIP strategies provides an inventory of which specific strategies were most easily adopted and which were used least, the longitudinal analysis that follows provides a more sophisticated overall analysis of teachers' use of the CISIP model over not just one, but 2 years of PD.

Research Question #2: Predictors of Teachers' PD Implementation

To refine our analysis, we designed two two-level HLMs. Both models were compared against a null model, i.e., a model with no predictors at either level of the analysis. This was to ensure there was variance to model at each level by the predictors we would ultimately include. It would also provide a baseline fit statistic with which to compare more complicated models. We used the total raw DiISC measures to describe teacher characteristics that might predict teachers' levels of implementation of a scientific classroom discourse community in their own classrooms. Of note is the fact that while no individual student-level information was available, we used the percentage of each teacher's school's students who qualified for a free and reduced lunch program. Also, we used the variables to describe potential factors that may account for change over time in the amount of PD strategies the teachers used. The two models, Model A and Model B, are described in the following equations:

$$\text{Model A} \quad \text{Level 1}: \text{PD Use} = \Pi_0 + \Pi_1 * (\text{time}) + e$$

$$\text{Level 2}: \quad \Pi_0 = \beta_{00} + \beta_{01} * (\text{SES}) + r_0$$

$$\Pi_1 = \beta_{10} + \beta_{11} * (\text{experimental condition}) + r_1$$

$$\text{Model B} \quad \text{Level 1}: \text{PDUse} = \Pi_0 + \Pi_1 * (\text{time}) + e$$

$$\text{Level 2}: \quad \Pi_0 = \beta_{00} + \beta_{01} * (\text{SES}) + r_0$$

$$\Pi_1 = \beta_{10} + \beta_{11} * (\text{total PD participation}) + r_1$$

We systematically tried every available predictor. The two resulting models were the only combinations of predictors that predicted with statistical significance. Both models fit similarly well,[1] having, statistically significant predictors for intercept and slope. However, the actual predictors of slope differed; that is, they were different conceptualizations of treatment. In Model A, treatment was a simple 1 or 0 grouping value. In Model B, that group membership was reflected by the actual amount of PD that any one teacher received. The HLM approach allowed for participants to enter or leave the PD and have different total amounts of participation in the program at any point in time. Because we were unable to analytically choose Model A or B (i.e., there was no statistically significant difference in fit), and both models indicated the same treatment effect (i.e., there was no qualitative differences in inferring an effect of treatment), we defer to discussing both models in making inferences about that treatment effect (see Tables 7 and 8 for the estimated parameters).

[1]Of note is that, unlike with traditional modeling techniques, we were unable to provide effect sizes. HLM requires that we consider model fit and only produces pseudoeffect sizes.

TABLE 7
Model A

	Effect (Variable)	*B*	*Se*	*t* Ratio	*df*	*p* Value
Intercept, Π_0	Intercept, β_{00}	38.22	4.70	8.13	58	< 0.01
	Poverty	−19.48	7.96	−2.45	58	0.018
Slope, Π_1	Intercept, β_{10}	−0.012635	0.011984	−1.05	58	0.297
	Condition, β_{11}	0.023016	0.009537	2.41	58	0.019

TABLE 8
Model B

	Effect (Variable)	*B*	*Se*	*t* Ratio	*df*	*p* Value
Intercept, Π_0	Intercept, β_{00}	36.790893	5.046069	7.291	58	< 0.01
	Poverty	−18.719641	8.550526	−2.189	58	0.032
Slope, Π_1	Intercept, β_{10}	0.002363	0.009015	0.262	58	0.794
	Condition, β_{11}	0.000481	0.000198	2.431	58	0.018

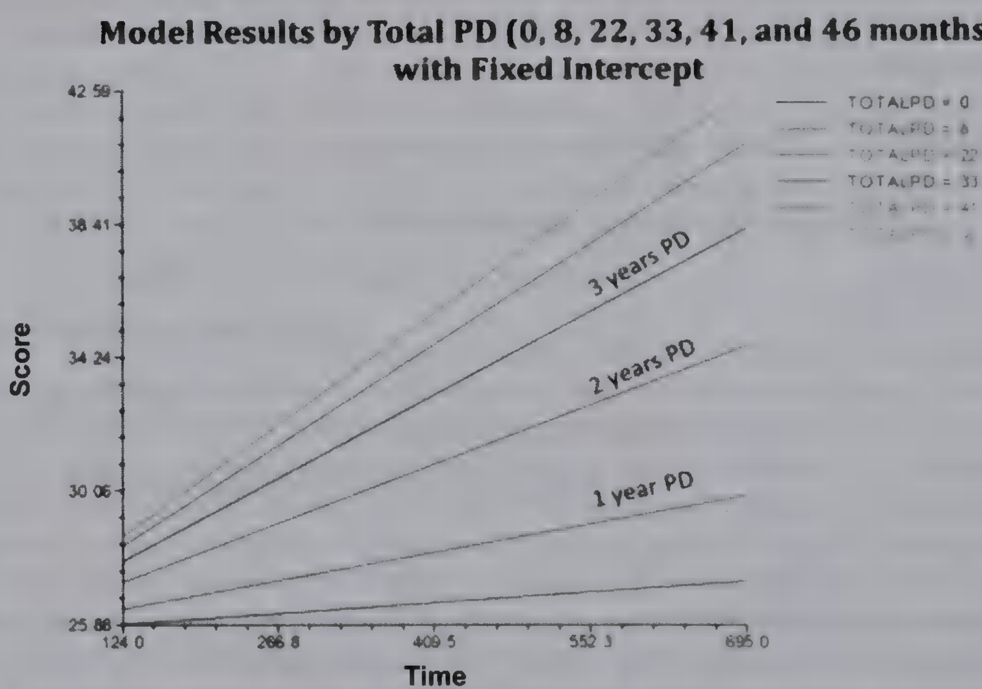


Figure 3. Slopes of teacher change due to amount of PD, holding intercept constant at zero. The lowest regression line represents the comparison group with no PD with an additional year of PD for each higher line.

In either Model A or B, the amount of PD, was the only statistically significant predictor of teachers' changing instructional strategies over time. Specifically, the more PD a teacher received, the more they used PD-corresponding instructional strategies. To illustrate, in Figure 3 the slopes of the lines represent the rate of change by different groups of teachers according to the amount of PD teachers received. In Figure 4, we hold the intercept constant and consider treatment as only a 0 (no PD) or 1 (PD) condition, simplified in the graph as follows: In either model, socioeconomic status (SES) was the only predictor of teachers' beginning use of PD-related strategies. Holding the slopes constant, we obtained the graph in Figure 5 to demonstrate differences in initial levels of PD. For the sake of completeness,

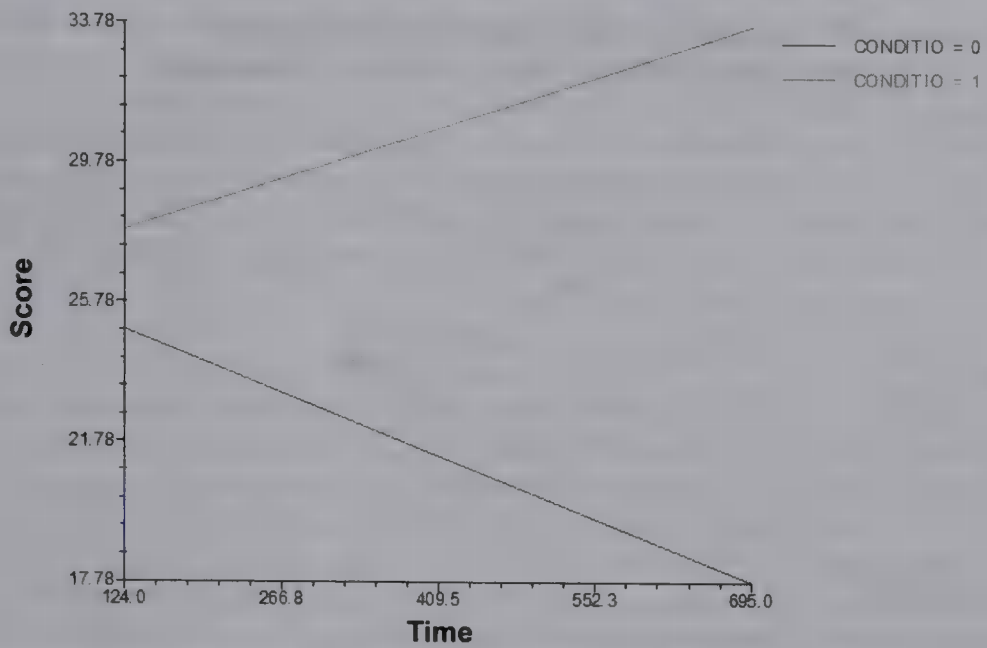


Figure 4. Teacher change slopes over time with and without PD.

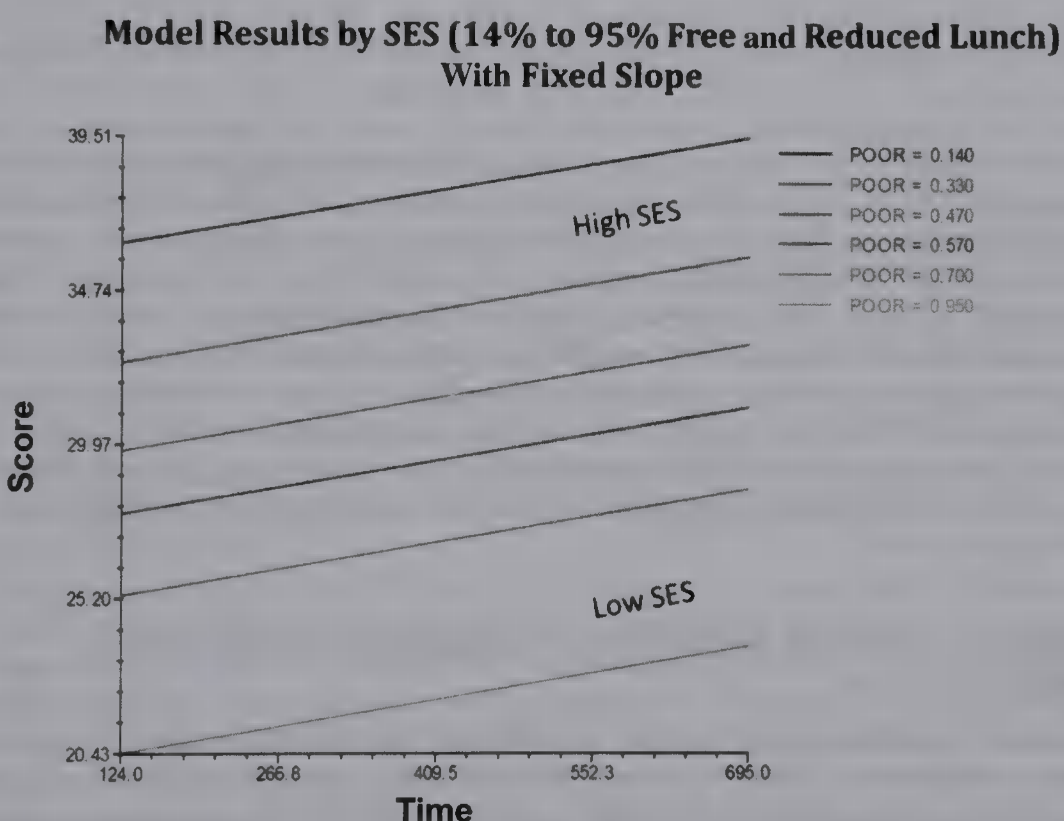


Figure 5. PD-related change over time holding slope constant. The percentage of students qualifying for free and reduced lunch decreases from 95% on the lowest regression line to 14% at the highest.

we include Figure 6, which allows both slope and initial SES to vary simultaneously, but it is complicated and thus we present further analysis of what the models mean in terms of teacher change.

We claim an effect on teachers' instructional practices, presumably due to the PD, as this effect was supported by both models' results and corresponding interpretations. This can

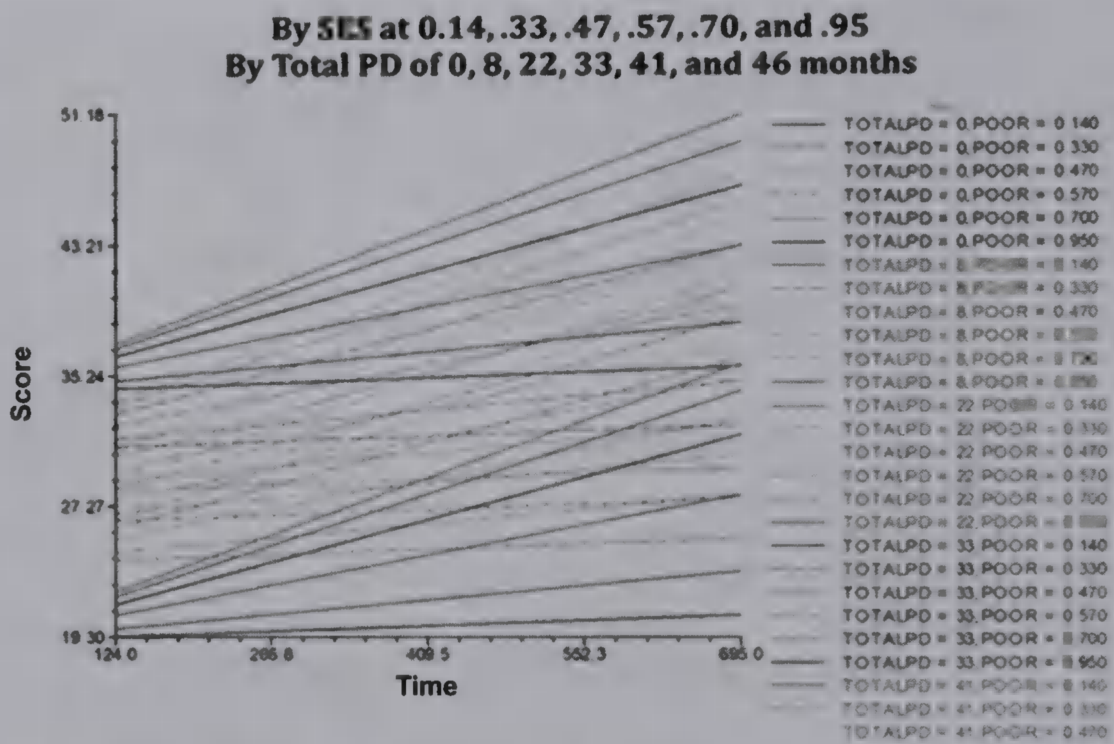


Figure 6. Complex full model that allows both slope and initial intercept (SES) to vary within subgroups.

be seen in Figure 3, where the intercepts, the teachers' starting points, were constrained to demonstrate how the slopes varied across levels of treatment, and in Figure 4, where only the group membership (with or without PD) was allowed to vary. While teachers increased in their use of CISIP instructional strategies, they began at a range of scores reflecting the average SES of their students. Figure 5 is a simplified graph of Model A where slopes were constrained according to specific levels of SES and treatment condition to demonstrate how the starting points of teachers varied across levels of SES. Figure 6 allows both SES and total amount of PD to vary simultaneously. In every graph, the effect of SES is uniformly related to the amount of initial, CISIP-related instructional practices that teachers used and the amount of PD (or whether they received it at all or not) determined use of PD-related strategies over time.

Research Question #3: Teachers' Prior Knowledge and Motivation to Change Instruction

Teachers' Experience, Certifications, and Subject Matter Knowledge. Our demographic survey results of science teachers' prior knowledge (i.e., educational background, preparation programs, and coursework) are presented in Table 3. Overall, there was a balance of new, midcareer, and veteran teachers with a variety of perspectives and experiences. Teachers were mainly in-field, secondary certified through either undergraduate or postbaccalaureate pathways. Nearly half of the teachers lacked a history and philosophy of science course. This lack of formal education in NOS, along with the observation data of science lessons in which teachers only occasionally engaged their students in discussions about NOS in conjunction with the science concepts they were studying, suggested that these teachers would benefit from learning more about NOS throughout the PD. Additionally, science teachers lacked expertise in English language arts content and associated teaching methods coursework in the use of written discourse and academic language development.

With a statewide requirement that teachers carry ELL endorsements, it was not surprising that most had had at least one ELL methods class. However, when we observed the science teachers such language-based instructional strategies were not often used. This suggests that all teachers needed even more opportunities to discuss and practice ALD and discourse, particularly written strategies. Despite our efforts to determine a pattern of which prior knowledge variables might predispose teachers to more readily adopt the PD model, none of these variables proved to be significant in our modeling process, nor in our general inspection of the data.

Teachers' Desire and Motivation to Change Instruction. To better understand teachers' level of motivation to change their teaching practices through the PD, we administered the *CISIP Teacher Self-Reflection Survey.* Based upon the sign tests, all differences between teachers' current and desired median teaching practices were significant at the $p < .001$ level, except for item #19, "How often during the week do students get information through lectures?," which was significantly different, but at the $p < .05$ level (Table 9). Six survey items (#7, 10, 14, 15, 17, and 20) required teachers to self-assess how often students engaged in inquiry-based instruction and activities; on average, the teachers rated their desired practice to more frequently include these inquiry-based instructional strategies than their current use. For instance, teachers wanted their students to develop and recognize alternative explanations for data, construct their own understanding of scientific concepts, defend their ideas with scientific evidence, and engage in hands-on activities more often. However, while teachers expressed the desire to change, based on our classroom observations of their teaching they still struggled with more frequent implementation.

Four survey items (#2, 3, 5, and 8) concerned oral discourse strategies and opportunities for students to talk with each other. Teachers wanted to include more student presentations, peer-to-peer discussions of their data, and whole-class discussions about NOS. Two items on the survey (#1 and 9) asked teachers to determine how often students engaged in writing-related activities. Results indicated that teachers desired to increase how often they had their students write about scientific investigations and revise their scientific writing ($z = -5.10$, $p < .001$). Again, while teachers reported that they wanted to use more oral discourse, the classroom results were mixed; some oral discourse strategies appear to be more easily integrated into teachers' instruction, whereas more formal aspects of scientific writing were less frequently used.

Two items (#6 and 11) inquired about specific strategies to increase students' academic language comprehension, having "students relate subject matter to their own experiences in other subjects or their own personal lives," and "acquiring scientific vocabulary through alternative means such as visual and/or kinesthetic activities." Teachers reported that they also wanted to increase how often they used these strategies. In practice, when we observed lessons, we sometimes saw the more easily adopted ALD strategies having to do with vocabulary acquisition and visual aids, but rarely saw differentiated instruction based on students' language capabilities or teachers explicitly bridging students' language and culture with the academic language and culture of science.

Teachers also indicated that they wanted to use learning principles more consistently in their classrooms. Five items (#4, 12, 13, 16, and 18) concerned opportunities for students to engage in various activities such as accessing prior knowledge, constructing conceptual frameworks, and engaging in metacognitive practices. Providing students with feedback on their written work is also in this category (item #18). Item #12 ($z = -4.903$, $p < .001$), addressing students' abilities to plan and organize their learning as an aspect of executive control and metacognition, and #13 ($z = -4.903$, $p < .001$), addressing students' writing

TABLE 9
Results of Sign Tests on Science Teachers' Responses to the "CISIP Teacher Self-Reflection" Survey Items

Item Pair (Current-Desired Use)	Negative Difference	Positive Difference	Number of Ties	Z	Significance
How often do students . . . ?					
1. write about science investigations	0	21	9		.000[a]
2. share findings through presentations	0	24	6		.000[a]
3. discuss data and understanding of meaning with peers	0	27	3	–5.004	.000
4. write and/or discuss ideas about concepts to be studied	0	25	5		.000[a]
5. engage in whole class discussions about the NOS	0	22	8		.000[a]
6. relate subject matter to their own experiences or lives	0	23	7		.000[a]
7. develop and recognize alternative explanations for data*	0	26	3	–4.903	.000
8. engage in discussion to acquire language structure and vocabulary appropriate for science communication	0	26	4	–4.903	.000
9. revise their writing about science	0	28	2	–5.103	.000
10. construct their own understanding of scientific concepts through observation and writing their own definitions	0	25	5		.000[a]
11. acquire scientific vocabulary through alternative means (visual and/or kinesthetic activities)	0	23	7		.000[a]
12. plan and organize their learning*	0	26	3	–4.903	.000
13. write/discuss before, during, and after a unit of study to identify their changing ideas and how they arrived at them	0	26	3	–4.903	.000
14. defend their ideas with scientific evidence/data through discussion and writing	0	23	7		.000[a]
15. write and discuss their imaginative ideas as a means of exploring science phenomenon	0	26	4	–4.903	.000
16. discuss or write what they have learned after a science lesson	0	24	6		.000[a]
17. discuss how theories have the explanatory power to generate many testable hypotheses	0	25	5		.000[a]
18. receive feedback from you on their written work	0	20	9		.000[a]
19. get information through lectures	6	0	24		.031[a]
20. engage in hands-on inquiry activities	0	18	12		.000[a]

Significance indicates a difference between the medians of the paired measures.
[a] = Binomial distribution used.

and/or discussing "before, during, and after a unit of study to identify their changing ideas and how they arrived at these ideas about science" (i.e., metacognition) were also significantly different than the teachers' self-assessment of their pre-PD instruction. From our analysis of the most and least frequently used strategies, teachers more easily adopted CISIP strategies such as (a) establishing community norms in the classroom, (b) providing clear feedback and teacher expectations, and (c) metacognitive opportunities (although less often). Throughout the first year of the PD, teachers struggled to change their instruction to become more reliant on assessing students' prior knowledge, helping students to become more self-aware and self-monitoring of their learning, and providing opportunities for students to have executive control of their learning.

Summary: Greatest Desired Areas of Change. The top five (25%) strategies that the teachers identified as their most desired changes were to have (a) #13, "Students write and/or discuss before, during and after a unit of study to identify their changing ideas and how they arrived at these new ideas about science" (+1.90); (b) #12, "Students plan and organize their learning" (+1.77); (c) #15, "Students write and discuss their imaginative ideas as a means of exploring science phenomenon" (+1.63); (d) #17, "Students discuss how theories have the explanatory power to generate many testable hypotheses" (+1.60); and (e) #9, "Students revise their writing about science and in particular, their own investigations" (+1.53). Thus, in theory these five instructional strategies could be targeted as ones that teachers would be initially most receptive to learning and implementing in their classrooms. In our observations, we saw that teachers made changes within one year of engaging with PD by providing more opportunities for peer-to-peer oral discourse and pre- and informal writing within the context of guided inquiry activities. They demonstrated less change in providing opportunities for student-designed inquiry investigations and executive control of learning. Teachers also appeared to need more encouragement and practice to integrate opportunities for students to learn about NOS, which could have addressed their desire for students to better understand hypotheses and theories within science.

Research Question #4: Teachers' Views of Barriers and Supports to PD Implementation

Using our survey items, teachers assessed perceived barriers and supports for implementing what they learned during the CISIP program. Teachers identified more sources of support than barriers; however, we did not ask them to weight each factor and we acknowledge that even one negative factor may be sufficient to prevent teachers from implementing what they learn through PD. Table 10 summarizes the percentage of items in each area that middle and high school teachers identified as barriers to, and supports for, PD implementation. Overall, comparable percentages of barriers and supports were identified by middle and high school science teachers. On average, the high school science teachers rated 23 items (51%) on the survey as a minor or major support, 18 items (40%) as neither a support nor a barrier, and only four items (9%) as barriers, but all barrier items identified by these teachers concerned parents and students. Middle school science teachers rated 21 survey items (47%) as supports, 16 items (36%) as neither, and eight items (18%) as barriers. Five of the eight barriers (63%) identified by these teachers concerned parents and students, whereas the other three included standardized testing, class size, and teacher team meeting and planning time. Overall, more items were considered supports (3.5 or greater) than barriers (2.5 or less) (Figure 7). Owing to space limitations only

TABLE 10
Summary of Supports, Barriers to PD Implementation (or Neutral) for Middle and High School Teachers

	Middle School			High School		
Area	Supports (%)	Neutral (%)	Barriers (%)	Supports (%)	Neutral (%)	Barriers (%)
All areas ($n = 45$ items)	47	36	18	51	40	9
Administration ($n = 4$)	100	0	0	100	0	0
Collaboration ($n = 9$)	44	44	11	66	34	0
Curriculum ($n = 5$)	40	60	0	20	80	0
Instruction ($n = 17$)	59	29	12	65	35	0
Students ($n = 8$)	13	37	50	13	50	3
Parents ($n = 2$)	0	50	50	0	50	1

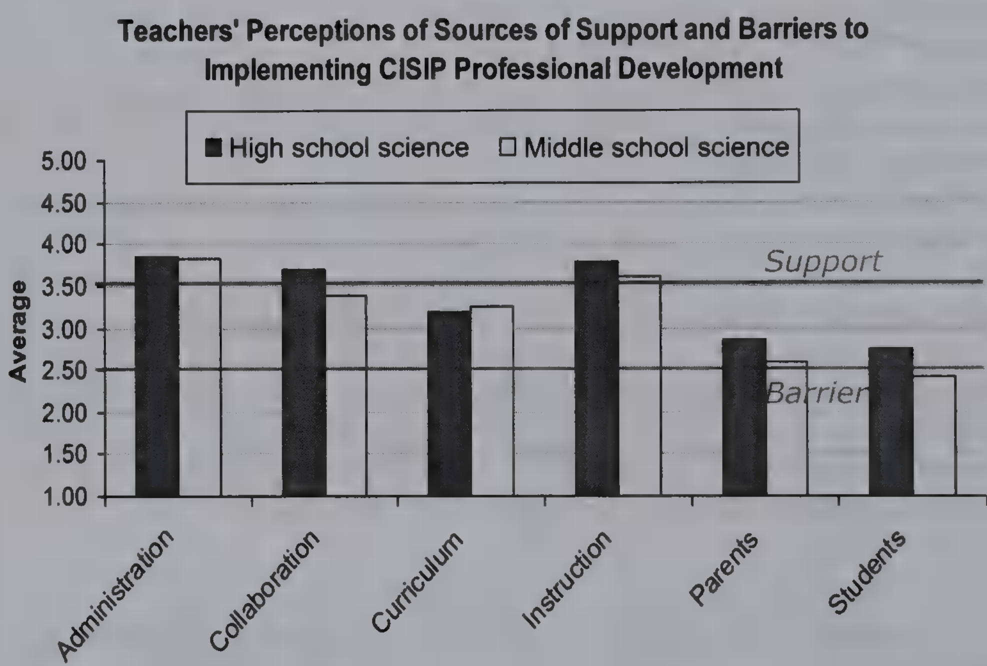


Figure 7. Graph of average responses by barriers and supports survey item categories.

those factors that were considered to be barriers to implementing the CISIP model are discussed.

Parents. Middle and high school science teachers perceived parents' attitudes toward the CISIP curriculum as neutral ($M = 3.18$–3.29). However, both the high school ($M = 2.43$) and middle school ($M = 2.00$) science teachers saw parents' ability to help their students with writing and discourse as a minor barrier to implementing the SCDC model. Whether or not these perceptions were accurate, teachers' beliefs could affect the amount and level of homework assignments that teachers gave to their students. If parents were viewed as being able to help their children at home, teachers might assign more challenging tasks, but if home support was perceived as absent, little or no homework might be assigned. Even the types of assignments that would be started in class and then need to be finished at home might be limited in scope.

Students. Middle school science teachers viewed students somewhat more negatively than high school teachers, on average identifying four items as minor barriers to implementation, as opposed to three items. Both high school ($M = 2.43$) and middle school ($M = 1.91$) science teachers perceived students' diverse language skills as minor barriers to CISIP implementation. Both also identified their students' grade-level background knowledge and writing and discussion skills as a minor barrier. Finally, middle school science teachers ($M = 2.18$) identified their students' attendance as a minor barrier to implementing CISIP.

DISCUSSION

We sought to document and investigate the following aspects of teachers' learning through the CISIP program: (a) prior education, teacher certification, and length of teaching experiences; (b) desire to change current teaching practices to be more aligned with the CISIP model; (c) use of specific PD strategies initially used within 1 year of PD and the overall change in their teaching practice over 2 years; and (d) identification of those factors that were supports or barriers to implementing the PD. When we synthesize the results in light of the PD, we see several trends: (a) teachers who are better able to engage their students with the nature of scientific communication, (b) the benefits of iterative PD with a complex task such as teaching science, and (c) the challenges of changing teachers' beliefs about how people learn and enacted instructional practices to match their desire for reform in the classroom. We discuss the relevance of these findings here in a broader context.

Teachers' Professional Development Concerning the Nature of Scientific Communication

The CISIP community of practice included a range of teachers that provided a balanced distribution of new, midcareer, and veteran teachers with a variety of perspectives and experiences. When we observed teachers' science lessons, we noticed that they only occasionally engaged their students in discussions about NOS. That nearly half of the teachers lacked a course in the history and philosophy of science suggested that most, even experienced, teachers would benefit by learning more about NOS as they developed and implemented science lessons. Since the 1990s, science education reform documents (NSES, NRC, 1996; Achieve, 2013) have encouraged the use of authentic learning experiences that reflect the ways in which scientists undertake and communicate their own work. Scientists work in teams of researchers, peer-review each other's work, and communicate their findings through a variety of oral and written modes. Thus, to better reflect the practice of doing science in authentic ways, all science teachers need to be able to bridge academic language and practices of scientists with students' everyday language and conceptions of the world around them.

The CISIP program was designed to help science teachers develop greater expertise and skills to implement instructional strategies in writing and academic language development to support students' learning of science. With less formal education in language arts and literacy strategies, the science teachers were less likely to integrate written discourse into their science lessons. In this study, we found that teachers, even with explicit PD activities on how to integrate writing into their science lessons, rarely engaged students in formal scientific writing or provided rubrics for their students to revise their writing. Science teachers also rarely provided differentiation in instruction or found ways to bridge language and culture with science. These sorts of communication and critical thinking skills are vital to a well-rounded education and have been carefully delineated in the new national

science education standards (Achieve, 2013). The NGSS also include cross-references to the Common Core English language arts standards that further emphasize the critical role of language in learning science and developing scientific literacy.

Throughout the CISIP program, teachers did improve in their use of small-group discussion even though they still relied upon whole-group classroom instruction. This improvement reflects a move toward adopting more aspects of a scientific classroom discourse community while still retaining teacher control, but it was a noticeable shift in teachers' practices. Lemke's (1990) identification of classroom triadic dialogue (IRE) as a means for knowledge transmission and discourse structure is the antithesis of science education reform as it prevents students from sharing control of the classroom discourse. However, Lemke found that it is a favored staple of whole-group discussion pedagogy in science classes. The CISIP program provided examples of how to shift the discourse in the classroom to establish more equitable and interesting learning opportunities for students. The use of social constructivist scientific inquiry as a teaching paradigm provides students with more opportunities, not only to engage with scientific questions, make observations, and make meaning from their own experiences, but also to talk with each other and not just their teacher. Gee (2004) argued that students need such peer-to-peer learning experiences to create meaningful discourse and develop conceptual understandings. Kelly (2014) identified discourse as one of the emerging research directions in science education in his review of discourse practices.

As shown in our model of instructional change, by trying to engaging students in SCDCs, CISIP teachers made some progress in changing their instruction to be more aligned with Vygotsky's sociocultural theory of learning (1986) and constructivist tenants of inquiry-based teaching. The inclusion of both oral and written discourse also aligns with the second core NRC (2005) learning principle, the essential role of factual knowledge and conceptual understandings. However, the fact that these science teachers lacked prior knowledge of how to use these types of instructional strategies before the PD seminars indicates that teacher education programs themselves should consider how to prepare teachers to be able to better meet state and national science education standards. Sadler (2006) addressed this issue specifically in a science methods course in which there was a focus on argumentation, but found that preservice teachers rarely had an opportunity to try this in their student teaching placements. Thus, we have a self-perpetuating problem of a lack of oral and written discourse in science classrooms and a lack of modeling these scientific practices for future science teachers.

Legitimate Peripheral Participation: The Benefits of Iterative Professional Development

There was no significant difference between the middle and high school science teachers' use of new strategies learned at the CISIP seminar activities. This indicated that although the teachers participated in separate 3-week summer institutes, it did not measurably affect their implementation of the CISIP instructional strategies as new participants. However, there was a significant difference between previous and new participant groups in their use of the CISIP strategies. The previous participants had higher implementation scores. This suggests that a second iteration of the same PD program supported greater implementation by those who elected to stay with the program.

Initially teachers made small changes that did not require radical reengineering of how they managed their classrooms; the most readily adopted strategies were related to teacher-centered instruction and the least adopted were ones that would be found in more student-centered classrooms. A fully realized SCDC would be a classroom in which students were

empowered to generate questions for investigation, had access to resources to support their learning, and had structured opportunities to reflect upon their own learning to develop executive control and self-monitoring capabilities as lifelong learners. However, teachers also tended not to use much formative assessment to guide their instructional decisions, which indicates that they needed more explicit PD in how to be more responsive to students' learning needs (Black & Wiliam, 1998). Thus, it appears that the larger issue was that many of these science teachers resisted releasing control and struggled with generating more opportunities for student choice and self-regulation. The easiest paths to new types of instruction were taken first; those that were more difficult, more central to teachers' beliefs about effective science instruction, presented greater institutional and social friction and required more PD.

Teacher Change Over Time

The length of time that the teachers received PD, or their experimental group membership, was chosen as the predictor of teacher change whereas a schools' percentage of students who qualified for free and reduced lunch was chosen as the exclusive predictor of the intercept or starting point. Over 2 years, the teachers who had participated for longer periods of time used more of the CISIP model strategies and had higher rates of change than newly participating teachers. The model indicated, with statistical significance, that SES predicted teachers' initial levels of PD-associated behavior. While the overall SES of the school's students was important in determining where teachers began, the amount of PD accounted for how teachers changed over time. When these commonsense results were supplemented with additional survey data, teacher beliefs were shown to be a dominating force. Specifically, survey data suggested that teachers believed students were the nearly singular barrier to implementing the CISIP model.

That said, the claim that CISIP was effective in changing teachers' practices was supported by both models. Because the same conclusion could have been drawn from both models about the effect of treatment, despite the different ways of coding treatment or nontreatment group membership, we concluded that the results did not depend on the coding system we used, but rather reflected a measurable change in teacher instruction. Our conclusion has several caveats. First, consider the multileveled regression lines in Figure 4. On a long enough timeline, the comparison group teachers' PD-associated behaviors would become negative and the CISIP teachers' PD-associated behaviors would approach infinity. But the CISIP measure has no meaningful negative or very large values. The linear nature of the relationship, outside the range of our data, was de facto absurd. This indicated that, although our models fit tolerably well, such a fit would not apply outside the range of our data. That is, we do not know whether increases in CISIP-related instructional practices over time will continue or drop off. For example, Rogers (2003) found that PD that requires less fidelity is more likely to be sustained over time. The CISIP model of a SCDC is complicated with a high cognitive load and appears to require multiple iterations to increase fidelity and teacher change, but did allow for teacher choice. Second, our final models suggested that initial implementation of PD was positively influenced by the average SES of the teachers' students with lower implementation associated with lower SES schools and higher implementation with higher SES schools. Our findings, produced using modern statistical methods, support the work of Anyon (1981) and Oakes and Guiton (1995) in that tracked, low-SES students in this study were initially taught with little or no inquiry-based science instruction. Statistically significant variance components, however, led us to believe that there might be other hitherto, unidentified factors that influenced the initial implementation of the CISIP strategies. Other possible factors could include teacher beliefs, systemic

barriers, school culture in terms of what constitutes good teaching, high-stakes testing, community expectations, the number of early adopters in the school, and the cognitive complexity of the CISIP instructional strategies.

Teachers' Desire to Change Instructional Practices

As they began CISIP, teachers expressed a desire to increase the frequency of how often students engage in behaviors that reflect a rich SCDC. For instance, teachers wanted their students to develop and recognize alternative explanations for data, construct their own understanding of scientific concepts, defend their ideas with scientific evidence, and engage in hands-on activities more frequently. Teachers also wanted to include more student presentations, peer-to-peer discussions of their data, and whole-class discussions about NOS than they currently did. The science teachers also had a statistically significant desire to increase the frequency of having their students write about scientific investigations and revise their writing about science and their investigations. This suggests that the desire to change is often strong, but the reasons for teachers' actual change or resistance to change requires more information about what supports and prevents such changes as a result of PD.

Barrier and Supports to Implementing Learning From Professional Development

In general, CISIP teachers identified administrative support, the PD strategies, and teacher collaboration as strong support for implementing new instructional methods. The lead CISIP designer had been a state science specialist and recruited teachers from districts that already had the support of the administration for the types of changes that the PD had proposed; thus it is not surprising, but rather validating, that the teachers identified administrative support of change. That the teachers also identified the PD strategies themselves as supportive reinforces the CISIP design as a viable model of an SCDC. Finally, the fact that teachers identified positive collaboration as a support for implementing new ideas underscores the value of engaging teachers in a community of practice. However, the teachers viewed students' grade-level science knowledge, diverse language skills, and discourse abilities as the greatest barriers. We recognize that teachers' beliefs and decisions about what and how to teach are complex and that in the future survey items may need to be weighted in terms of how critical teachers view each factor.

It is problematic that CISIP teachers perceived their students to be a barrier to using what they learned in PD. When teachers, especially those teaching lower tracked students or students in working-class communities (Anyon, 1981; Lee et al., 2013), believe that students are unable and/or unwilling to engage in critical thinking and inquiry-based science investigations, they fail to provide such opportunities, thus limiting students' access to a standards-based science education (Oakes, 1995). Even for experienced teachers who are past the induction phase of teaching and are confident in their teaching abilities, PD may need to explicitly address teachers' dispositions toward equity in the classroom (Kelly, 2014).

Professional Development Interaction With Policy and Politics

Van Driel, Beijaard, and Verloop (2001) emphasized the value of teachers' practical knowledge as experts in their own classrooms and recommended engaging teachers in long-term staff development so that teachers have time to restructure their knowledge and beliefs and integrate new information with their practical knowledge. National priorities for science, technology, engineering, and mathematics (STEM) education and recruiting

students into STEM careers have been outlined in numerous policy documents (e.g., Committee on Prospering in the Global Economy of the 21st Century, 2007). When U.S. National Science Education Standards were introduced, there were clear goals for reformed science teaching to use constructivist inquiry-based instruction to foster more robust learning opportunities for students (NRC, 1996), thus preparing them to be scientifically literate citizens and perhaps productive STEM professionals. A decade later, these recommendations had become increasingly difficult to address within the pressures of high-stakes testing (Nichols & Berliner, 2007). These challenges require several responses if we desire educational reform. First, administrative support for science teacher PD needs to permeate schools and districts so that there is institutional momentum that supports teacher change. Administrative support is critical so that teachers know that they will be supported when they adopt new instructional methods (Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2009). Second, and perhaps more importantly, PD itself must empower teachers to change their instruction and institutions to change their attitudes toward teacher instruction. PD cannot only be for teachers, but must be for entire organizations, requiring changes to support, or hinder, behavior that helps, or hurts, students.

Methodological Limitations

The effect of this instance of teacher PD and our inferences about it were limited by the methodology in four primary ways. First, we were unable to evaluate treatment infidelities. Teachers may have attended the PD sessions, but there are multiple levels of engagement and each teacher has their own unique learning experience. How teachers translated the CISIP program concepts and instructional strategies into their classrooms was evaluated, but the extent to which it systemically changed their instruction, perhaps even permanently, was unknown. In fact, our assumption was associated with the teachers' interaction with the PD, which limited our generalizations to other groups of teachers.

Second, our sample was one of convenience; teachers were not randomly selected to be or not be in our study, nor were they randomly assigned to treatment or control (nontreatment) groups. This limited our inferences by the sampling procedure we employed. Third, we must limit our inferences to the boundaries of our data. In the same way that we cannot make judgments about teachers' change over the first year of their PD, we cannot make inferences about changes due to more sustained PD beyond the 2-year study. Research has shown that sustained PD creates lasting effects (Blank et al., 2008), but we were unable to verify those results in that our research and the PD itself was limited to the project's funding.

Lastly, and perhaps most importantly, we unable to construct sufficient validity and reliability arguments for the DiISC to ensure that measurement error itself did not limit our inferences. We are able to make claims about the effect of the PD, but only an effect as demonstrated on our measure of treatment implementation. Other, more distal measures would require further study and more complete validity and reliability arguments associated with our outcome measures.

In summary, while we were able to make inferences, we were unable to make the broad generalizations we would have liked. Specifically, we were limited by an inability to assess systemic changes in teacher practices or to infer beyond the boundaries of our data, especially with respect to sustained PD over longer periods of time and to other studies that also used our DiISC instrument. Future studies should involve a reliability and validity argument sufficient to make such generalizations, better measures of systemic teacher change, and should take advantage of the possibility of extended data collection or traditional random assignment and its advantages.

CONCLUSIONS

This study investigated changes in teachers' science instruction as they progressed through a particular, iterative teacher PD program. Within the categories that were used by van Driel et al. (2012), this study would be classified as one that explored the relationships among the external domain, domain of practice, and the personal domain. A framework of cognition, beliefs, and situated learning allowed us to analyze teachers' perspectives and change to make limited inferences. Like the CISIP model itself, the members of the learning community demonstrated that "learning is not merely a condition for membership, but is itself an evolving form of membership" (Lave & Wenger, 1991, p. 53). Teachers who entered into the PD acquired, through oral and written discourse, practice, and collegiality, an initial understanding of how to build scientific classroom discourse communities.

The CISIP model is based on the concept of teacher learning communities as a means for affecting positive change for student learning within inquiry-based science instruction. By participating in the CISIP learning community, individuals increased their awareness of many different types of teaching strategies. Further research on SCDCs and other similar PD programs may not agree with the CISIP model, but it was the background for the data generated and analyzed in this study. The longitudinal model clearly indicated that the CISIP model of iterative PD works, although not without its challenges, to change teachers' instruction to incorporate more aspects of scientific communication, and we believe others like it will similarly work, drawing on the same learning principles and relationships previously described.

Teacher PD has equity and policy implications at the school, district, state and national levels. Thus, we make the following recommendations. First, that complex, change-inducing PD should be iterative. On average, teachers used more of the CISIP model as they engaged with it repeatedly over time. Initially, more easily changed teacher-centered strategies were adopted, followed by strategies that were student centered, as these actions required more radical departures from extant instruction. Second, facilitate teacher learning by structuring PD through legitimate peripheral participation and ZPD. In our study, as teachers became mentors and facilitators their use of the CISIP model was more sustained. Previous participants acted as formal and informal mentors to newer participants, and these more experienced CISIP teachers shared the results of trying new approaches in their own classrooms. Third, use explicit modeling and planning in PD activities to encourage implementation. For example, all the teachers in CISIP used science notebooks with their students to some degree. It was the most readily adopted piece of CISIP, and the PD activities were very clear regarding how to use notebooks with students. Fourth, striking a balance between presentation and practice of PD material at the workshop sessions, combined with planning time throughout the academic year, potentially increases teachers' levels of implementation. In our study, teachers valued planning time with their team members during the PD sessions. Finally, teacher perceptions and expectations of student learning must be challenged during PD. Teachers who had equitably high expectations for student learning were more open to using the CISIP model with all of their students, not just their high-performing students. Teachers who differentiated between students used more of the model, and more inquiry-based instruction, with the students they perceived as being generally more capable (e.g., college bound) in science; and these students usually had a higher SES. Ultimately, through PD, teachers must view all their students as capable of engaging in inquiry-based scientific thinking.

Science teacher PD providers can benefit educational reform movements by leveraging broader conceptions and frameworks of teaching and learning, such as a SCDC. External factors, e.g., school culture, can unwittingly block teachers from implementing new ideas.

In particular, the pressures on science teachers concerning low-performing students and state-mandated testing can have an unintended effect of derailing a teacher's efforts to enact equitable constructivist instruction. Thus, administrative support is critical to teacher change. Internal factors, such as teachers' beliefs about students' cognition, also have both great potential and danger to affect the range of learning opportunities made available to all students.

Recommendations for Future Research

Considering teachers' overall view of students as a primary barrier, and their parents as a secondary barrier, to implementing new instructional methods, researchers should investigate how PD might confront teachers' differential views of students' abilities and capability to engage in scientific inquiry. When implementing a new teaching approach, teachers in CISIP appeared to adapt it according to their own institutional contexts and beliefs. In future studies, contexts and beliefs would be important predictors to investigate more closely to better understand teacher change. Simple group membership and amount of PD, while important for understanding teacher change, needs to be expanded to encompass the factors within that PD and teacher learning that matter most.

In addition to Wilson's (2013) summary of needed research in teacher PD, we offer several recommendations for future research to construct longitudinal models of teachers' use of PD, change, and effectiveness: (a) account for discrepancies in sampling procedures to eliminate plausible, alternative hypotheses in search of causal links; (b) frequently observe teachers over long periods of time and long after the PD has ended; (c) make more frequent observations of teachers over time with a stronger understanding of baseline practices for a more precise chronicling of teacher change; (d) in making claims about the effects of PD, researchers must take care to ensure that the measures used to make those claims have adequately developed validity and reliability and/or credibility and transferability arguments; and (e) include student outcomes. Providing high-quality PD that results in teacher learning and implementation, as well as a positive effect on student learning outcomes, is an aspect of educational research that has been neglected. We cannot expect to improve schools, thus fulfilling their democratic mission through equitable student achievement, if teacher PD programs are not built upon both sound learning theories and reliable findings as to their effectiveness to reform science instruction.

REFERENCES

Achieve, Inc. (2013), on behalf of the twenty-six states and partners that collaborated on the NGSS. Next Generation Science Standards. Achieve, Inc.

American Association for the Advancement of Science (1993). Benchmarks for scientific America's public schools. New York: Basic Books.

Anyon, J. (1981). Social class and school knowledge. Curriculum Inquiry, 11(1), 3–42.

Baker, D., Beard, R., Bueno-Watts, N., Lewis, E., Özdemir, G., Perkins, G., Uysal, S., Wong-Kavas, S., & Yaşar-Purzer, S.* (2008). Discourse in Inquiry Science Classrooms (DiISC): Reference manual (Technical Report No. 001). The Communication in Science Inquiry Project (CISIP): Arizona State University. (*Development team authors listed in alphabetical order after first author.)

Baker, D. R., Lewis, E. B., Purzer, S., Watts, N. B., Uysal, S., Wong, S., Beard, R., & Lang, M. (2009). The Communication in Science Inquiry Project (CISIP): A project to enhance scientific literacy through the creation of science classroom discourse communities. International Journal of Environmental & Science Education, 4(3), 259–274.

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the National Science Foundation's Local Systemic Change through Teacher Enhancement initiative. Journal of Research in Science Teaching, 44(3), 375–395.

Bellanca, J., & Brandt, R. (2010). 21st century skills: Rethinking how students learn. Bloomington, IN: Solution Tree Press.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. The Phi Delta Kappan, 80 (2), 139–148.

Blank, R. K., De las Alas, N., & Smith, C. (2008). Does teacher professional development have effects on teaching and learning?: Analysis of evaluation findings from programs for mathematics and science teachers in 14 states. Washington, DC: Council of Chief State School Officers.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. Educational researcher, 33(8), 3–15.

Borko, H., & Putnam, R.T. (1996). Learning to teach. In D. C. Berliner and R. C. Calfee (Eds), Handbook of educational psychology (pp. 673–708). New York: Macmillan Library Reference, London: Prentice Hall International.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. Teaching and Teacher Education, 18(8), 947–967.

Committee on Prospering in the Global Economy of the 21st Century (US), Committee on Science, & Public Policy (US). (2007). Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington, DC: National Academy Press.

Cuban, L. (1992). Curriculum stability and change. In P. W. Jackson (Ed.), Handbook of research on curriculum (pp. 216–247). New York: Macmillan.

Darling-Hammond, L., & Bransford, J. (Eds.). (2007). Preparing teachers for a changing world: What teachers should learn and be able to do. San Francisco: Jossey-Bass.

Donovan, M. S. (2013). Generating improvement through research and development in education systems. Science, 340(6130), 317–319.

Dweck, C. S. (2000). Self-theories: Their role in motivation, personality, and development. London, Psychology Press.

Erickson, F. (1986). Qualitative methods in research on teaching (pp. 119–161). In M.C. Whittrock (Ed.) Handbook of research on teaching (2nd ed).

Franke, M. L, Kazemi, E., Carpenter, T., Battey, D., & Deneroff, V. (2002). Articulating and capturing generative growth: Implications for professional development. Paper presented at annual meeting of the American Educational Research Association: New Orleans, LA.

Fraser, B. J., Tobin, K., & McRobbie, C. J. (Eds.). (2011). Second international handbook of science education. Springer Science & Business Media.

Fraser, B. J., & Tobin, K. G. (2012). In C. J. McRobbie (Ed.), Second international handbook of science education. Dordrecht, The Netherlands: Springer.

Gee, J. P. (2004). Situated language and learning: A critique of traditional schooling. London: Psychology Press.

Gee, J. P. (2005). Language in the science classroom: Academic social languages as the heart of school-based literacy. In R. K. Yerrick & W.-M. Roth (Eds.), Establishing scientific classroom discourse communities: Multiple voices of teaching and learning research (pp. 19–44). Mahwah, NJ: Erlbaum.

Hand, B. M., Alvermann, D. E., Gee, J., Guzzetti, B. J., Norris, S. P., Phillips, L. M., et al. (2003). Message from the "island group": What is literacy in science literacy? Journal of Research in Science Teaching, 40(7), 607–615.

Hewson, P. W. (2007). Teacher professional development in science. In, S. K. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 1177–1203). Mahwah, NJ: Erlbaum.

Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. Journal of Research in Science Teaching, 42(6), 668–690.

Kelly, G. J. (2014). Discourse practices in science learning and teaching. In N. G. Lederman & S. K. Abell (Eds.), Handbook of research on science education (Vol. II, pp. 321-336). Mahwah, NJ: Erlbaum.

Kunzman, R. (2003). From teacher to student, the value of teacher education for experienced teachers. Journal of Teacher Education, 54(3), 241–253.

Lave, J., & Wenger, E. (1991). Situated learning. New York, Cambridge University Press.

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to Next Generation Science Standards and with implications for Common Core State Standards for English language arts and mathematics. Educational Researcher, 42(4), 223–233.

Lemke, J. L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Lewis, E. B. (2009). Secondary science teachers' view toward and classroom translation of sustained professional development. Doctoral dissertation, Arizona State University.

Lewis, E.B. (2011). Secondary Science Teachers' Translation of Professional Development through Affinity- and Institution-identity. Paper presented at the 2011 annual meeting of the National Association for Research in Science Teaching: Orlando, FL.

Lieberman, A. (1992). Introduction: The changing context of education. In A. Lieberman (Ed.), Literacy: Project 2061. New York.

Loucks-Horsley, S., Love, N., Stiles, K. E., & Mundry, S., Hewson, P.W. (2003). Designing professional development for teachers of science and mathematics (2nd ed.). Thousand Oaks, C A: Corwin Press.

Loucks-Horsley, S., Stiles, K. E., Mundry, M. S. E., Love, N. B., & Hewson, P. W. (2009). Designing professional development for teachers of science and mathematics. Thousand Oaks CA: Corwin Press.

Loughran, J. J. (2007). Science teacher as learner. In, S. K. Abell & N. G. Lederman (Eds.). Handbook of research on science education (pp. 1043–1065). Mahwah, NJ: Erlbaum.

National Research Council (1996). National Science Education Standards. Washington, DC: The National Academy Press.

National Research Council (2000). How people learn: Brain, mind, experience, and school. J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Washington, DC: The National Academy Press.

National Research Council (2005). How students learn: History, mathematics and science in the classroom, a targeted report for teachers. M. Donovan and J. Branford (Eds.). Washington, DC: The National Academy Press.

Nichols, S. L., & Berliner, D. C. (2007). Collateral damage: How high-stakes testing corrupts America's schools. Cambridge, MA: Harvard Education Press.

NGSS Lead States (2013). Next Generation Science Standards: For States, By States. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. Review of Educational Research, 78(1), 33–84.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking. American Educational Research Journal, 32(1), 3–33.

Özdemir, G., Lewis, E. B., Baker, D. R. (2007). Development and validity of the CISIP Classroom Observation Instrument (COI). Paper presented at the annual meeting of the National Association for Research in Science Teaching: New Orleans, LA.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. American Educational Research Journal, 44(4), 921–958.

Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? Educational Researcher, 29(1), 4–15.

Raudenbush, S., & Bryk, T. (2002). Hierarchical linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage.

Rogers, E. (2003). Diffusion of innovations. New York: Free Press.

Sadler, T. D. (2006). Promoting discourse and argumentation in science teacher education. Journal of Science Teacher Education, 17(4), 323–346.

Saul, E. W. (Ed.) (2004). Crossing borders in literacy and science instruction: Perspectives on theory and practice. Arlington, VA: NSTA Press.

Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

[State] Department of Education (2008). [State] Department of Education School Report Cards 2006–2007. Retrieved July 12, 2008.

Their, M., & Daviss, B. (2002). The new science literacy: Using language skills to help theory and practice. Arlington, VA: NSTA Press.

van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. Journal of Research in Science Teaching, 38(2), 137–158.

van Driel, J. H., Meirink, J. A., Van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. Studies in Science Education, 48(2), 129–160.

Vygotsky, L. (1986). Thought and language. Cambridge, MA: MIT Press.

Wallace, J., & Loughran, J. (2012). Science teacher learning. In Second international handbook of science education (pp. 295–306). Dordrecht, The Netherlands: Springer Netherlands.

Wilson, S. M. (2013). Professional development for science teachers. Science, 340(6130), 310–313.

Yerrick, R. K., & Roth, W. M. (Eds.). (2004). Establishing scientific classroom discourse communities: Multiple voices of teaching and learning research. London: Psychology Press.

Science
Education

Preservice Teachers Developing Coherent Inquiry Investigations in Elementary Astronomy

JULIA D. PLUMMER, ARZU TANIS OZCELIK
Department of Curriculum and Instruction, Pennsylvania State University, University Park, PA 16802, USA

Received 26 November 2013; accepted 14 April 2015
DOI 10.1002/sce.21180
Published online 15 June 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: For students to attain deep understanding of scientific practices, they will need to have opportunities to participate in sustained engagement in doing science. Such opportunities begin with elementary teachers implementing coherent and well-sequenced inquiry-based investigations in their classrooms. This study explored how preservice teachers ($N = 30$) planned inquiry-based investigations for elementary students. The preservice teachers spent the first 5 weeks of their methods course participating in astronomy investigations and then pair-taught astronomy investigations once a week for 5 weeks to elementary students in afterschool programs. We analyzed lesson plans, teaching reflections, and pre/post astronomy content assessments. One-third of the pairs developed coherent science inquiry investigations across all of their lessons. Their reflections suggest that preservice teachers who developed coherent inquiry investigations held normative ideas about scientific inquiry and were more likely to reflect on sense-making practices than preservice teachers who did not plan for coherent science inquiry investigations in their lessons. Preservice teachers' postinstruction astronomy content knowledge was positively correlated with an increased number of lessons spent on coherent science inquiry investigations. Based on our findings, we recommend engaging preservice teachers in coherent science inquiry investigations in a single domain followed by opportunities to plan and teach elementary children in that domain. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99**:932–957, 2015

Correspondence to: Julia D. Plummer; e-mail: jdp17@psu.edu
Supporting Information is available in the online issue at wileyonlinelibrary.com.

INTRODUCTION

Science education policy documents promote a view of classroom science in which students engage with core science ideas and scientific practices in ways that reflect what we know about how students learn most effectively (National Research Council [NRC], 2000, 2007, 2012). The *Next Generation Science Standards* (NGSS) raises the stakes for elementary science education. The NGSS suggests how science practices grow in complexity and sophistication across the grades, thus requiring extended engagement with the practices within and across grade levels as well as across content areas (NGSS Lead States, 2013). The NGSS places a particular focus on moving science education away from the fragmented and inconsistent "mile wide and an inch deep" approach common to U.S. K–12 education (Schmidt, McKnight, & Raizen, 1997) and toward a coherent integration of science content and the practices scientists use to engage in inquiry (NRC, 2012). For students to attain deep understanding of science in ways that reflect this integration of content and practice, teachers will need to provide opportunities for their students to engage with science in ways that help them see connections within science content and practice over time.

One method of supporting this type of learning is to organize instruction using a *coherent science content storyline* (CSCS; Roth & Garnier, 2006; Roth et al., 2011). Instruction built around a CSCS can "create a 'big picture' by purposefully selecting and sequencing science ideas in ways that build on one another" (Zembal-Saul, McNeill, & Hershberger, 2012, p. 48). In doing so, instruction moves away from disconnected, hands-on, science activities characteristic of the U.S. science classrooms (Roth et al., 2006) and toward instruction that has been shown to improve student learning (Roth et al., 2011).

We suggest that science teachers can provide students with a coherent experience that integrates science content and practices by combining elements of the CSCS framework with essential practices of scientific inquiry through a *coherent science inquiry investigation*. The term *investigation* identifies the unit of analysis under study, as an investigation can span multiple days or lessons. The term *science inquiry* then highlights what needs to be coherent within the investigation.

We consider a coherent experience in the practice of scientific inquiry as having the following characteristics: students have the opportunity to construct explanations based on evidence for scientific phenomena, in response to a scientific question or problem statement (NRC, 2000). Such an investigation may also engage students in additional science practices, such as the use of models and modeling, planning and carrying out an investigation, or communicating and justifying explanations to peers (NRC, 2000, 2012). However, in the interest of a unified analysis of whether or not preservice elementary teachers plan lessons around a coherent science inquiry investigation rather than disconnected science activities, we examined how they provided opportunities for students to engage in the *connections* between questions, evidence, and explanations.

The coherent science inquiry investigation framework parallels elements of the CSCS framework. A CSCS uses a goal statement or focus question to organize instruction; a coherent scientific investigation is also guided by a question or statement. In CSCS instruction, activities and content representations are matched to the learning goal and sequenced to support learning; in a coherent science inquiry investigation, activities and content representations are selected to help students answer the scientific question by gathering data and constructing an explanation, sequenced such that students are supported in making sense of evidence rather than confirming an answer provided by the teacher. Furthermore, both instructional frameworks focus on making appropriate connections and sequencing of content: for example, in a coherent scientific investigation, one investigation's

conclusion can lead to a new question that builds on previous findings and wonderings forming new connections across the content.

While previous studies have explored how teachers and preservice teachers adapt curriculum and engage elementary students in inquiry (e.g., Biggers & Forbes, 2012; Davis, 2006; Forbes, 2011, 2013; Forbes & Davis, 2010a; Hapgood, Magnusson, & Palinscar, 2004; Metz, 2004; Siry, Zeigler, & Max, 2012; Varelas et al., 2008), limited research has explored how preservice teachers create coherent science investigations for elementary students that attempt this integration of content and practice. Prior research suggests that elementary teachers favor the use of hands-on "activities that work," resulting in a conceptually fragmented science curriculum (Appleton, 2002, 2003); thus, more research is needed which explores the extent to which preservice teachers prioritize these coherent investigations over disconnected hands-on science activities and what methods might be employed to support them. We therefore examined how preservice teachers develop coherent science investigations in the context of their elementary science methods course as a step toward understanding how to help teachers implement instruction around core disciplinary ideas and science practices (NGSS Lead States, 2013; NRC, 2012) in ways that will support student learning through coherent science experiences (Roth et al., 2011).

ELEMENTARY TEACHERS' IDEAS ABOUT TEACHING SCIENCE AS INQUIRY

New elementary teachers (including preservice teachers and teachers with less than 5 years of experience) face several challenges in planning scientific investigations for their students. We first consider new elementary teachers' knowledge and beliefs about science inquiry. This is a complex construct to consider, as use of this construct draws on teachers' knowledge of the features of inquiry as well as their beliefs about how inquiry should be used in teaching. Yet, this distinction between teachers' knowledge and beliefs becomes blurry when attempting to analyze teacher thinking and action in classroom practice (Avraamidou & Zembal-Saul, 2010; Gess-Newsome, 1999). We therefore reviewed literature on new elementary teachers' ideas about inquiry and beliefs about inquiry-based pedagogy while also recognizing the difficulty in drawing this distinction.

New teachers' beliefs about the nature of science may lack sophistication and confuse science procedures with the nature of science (Brickhouse, 1990; Davis, Petish, & Smithey, 2006; Roehrig & Luft, 2004; Schneider & Plasman, 2011). Though some new teachers may describe inquiry as including questions and evidence, others believe hands-on activities and discovery-based approaches are science inquiry (Schneider & Plasman, 2011). However, Davis et al. (2006) suggest that most researchers focus on preservice teachers' understanding of specific skills rather than inquiry practices, which may limit the extent to which we can hypothesize about preservice teachers' ideas about coherent science inquiry investigations.

New teachers' beliefs and prior science education experiences may also influence the extent to which they engage their future students in scientific inquiry. In a study of two first-year elementary teachers, Avraamidou and Zembal-Saul (2010) describe the influence of teachers' knowledge of science teaching and beliefs on their practice. These promising beginning teachers engaged their students in the practices of science in ways consistent with their beliefs about science teaching. However, the two teachers differed in the extent to which they supported their students in sense-making practices of generating claims based on evidence. These differences in their knowledge of science teaching were also reflected in their personal beliefs about teaching science inquiry and may have been related to differences in their prior science learning experiences; one had taken extensive college-level

reformed-based science courses, whereas the other had only traditional science experiences prior to their science methods course.

New teachers may also face challenges in teaching science because of limited or incomplete content knowledge, affecting their choices when teaching science (Akerson, Morrison, & Roth-McDuffie, 2006), because they are typically not science majors (Gess-Newsome, 1999; Jones & Edmunds, 2006). Prior research suggests there is a significant positive relationship between elementary teachers' content knowledge and the extent to which they implement inquiry-based science lessons (Luera, Moyer, & Everett, 2005). Some elementary teachers have difficulty incorporating their content knowledge into their teaching, suggesting that their understanding of inquiry pedagogy is separate from their knowledge of the content (Alonzo, 2002). New teachers' limited content knowledge also points to the challenges they face with representing content appropriately for their students (Zembal-Saul, Blumenfeld, & Krajcik 2000), implying challenges for planning investigations. Content representations refer to teachers' knowledge of topic-specific instructional strategies. This can include analogies, models, demonstrations, and investigations. It can also include how they use an understanding of students' beliefs and abilities to structure and sequence interventions to appropriately address students' prior knowledge and their learning goals.

Researchers have also studied how new elementary teachers adapt science curricula toward addressing their personal knowledge and beliefs about teaching science. Though preservice teachers are capable of critical analysis of instructional materials, they do not often focus on how science concepts are represented and instead see inquiry as an opportunity to engage children's interest (Davis, 2006). Preservice elementary teachers are capable of shifting lesson plans to be more inquiry-focused, but the curriculum itself is also a mediating factor in the final inquiry orientation of adapted lessons (Forbes, 2011; Forbes & Davis, 2010a; Gunckel, 2011). Furthermore, the context in which the preservice elementary teacher enacts his or her lessons also influences the inquiry orientation of the lessons; these contextual factors may include school norms of curriculum material use, time constraints, and decisions made by preservice teachers' mentor teachers (Forbes, 2013).

SUPPORT FOR NEW ELEMENTARY TEACHERS ENGAGING STUDENTS IN SCIENCE INVESTIGATIONS

Prior research points to potential methods to ameliorate the challenges new teachers face in developing coherent science inquiry investigations. When preservice elementary teachers have the opportunity to take inquiry-based science content courses, they are better prepared to develop inquiry-based lessons than when they only take traditional college science courses (Avraamidou & Zembal-Saul, 2010; Haefner & Zembal-Saul, 2004; Luera et al., 2005). And when provided with a methods course that is structured to support their understanding of science practices, they are capable of adapting curricula in ways that provide students with opportunities to engage in science practices (Forbes, 2011; Forbes & Davis, 2010a). Forbes (2011) examined the ways in which preservice teachers adapted science lessons. The preservice teachers were able to develop investigation questions that drove the lessons, increase students' opportunities to gather data as evidence, and provide more structure for how to record and make sense of their data. The preservice teachers placed additional focus on supporting students' opportunities for constructing evidence-based explanations, such as including reflective journal prompts or by using models and representations to support their reasoning. These studies suggest that the extent to which preservice teachers' design inquiry-based lessons depends on the nature of their content

courses and the opportunities their methods courses afford to develop their understanding of science practices.

Providing preservice teachers with opportunities to engage in cycles of teaching and reflection may also support increased sophistication in their use of inquiry-based pedagogy. Zembal-Saul et al. (2000) investigated a pair of preservice teachers as they designed and delivered two cycles of planning, teaching, and reflecting on connected sets of three lessons for elementary students in their fieldwork placement. Providing structured opportunities to reflect on their teaching may have resulted in the improved engagement with science inquiry over time: In their second teaching cycle, the preservice teachers provided more opportunities for students to engage with appropriate phenomena firsthand and more opportunities for their students to engage in science practices, such as constructing explanations by identifying patterns in data. At the secondary level, Lotter, Singer, and Godley (2009) also found that preservice teachers' use of inquiry with their students improved after multiple cycles of teaching and reflection. The preservice teachers' views on inquiry and ability to engage their students in the practices of science increased during their second teaching opportunity. These improvements were a result of the guided reflections they made on prior teaching experiences. Reflection by preservice teachers can facilitate their pedagogical growth and make their beliefs visible when they are given proper support and scaffolding (Melville, Fazio, Bartley, & Jones, 2008).

These studies point to the importance of considering how new teachers' knowledge of science inquiry may shape their choices in planning inquiry experiences for their students. They also suggest that new teachers' understanding of the science content, the nature of the available science curriculum, the particulars of their teaching context, and opportunities to reflect on their teaching all influence how they develop science inquiry lessons. We drew upon these findings to design an elementary science teaching methods course to investigate a potential method of supporting preservice teachers in planning coherent science inquiry investigations and to examine the extent to which they planned for these coherent experiences over disconnected, hands-on activities. Thus, the following research question guided the study: How do preservice teachers' relevant science content knowledge, use of curricular resources, and understanding of scientific inquiry relate to their development of coherent science inquiry investigations?

METHODOLOGY

Study Design

This study used a mixed-methods design to investigate how an elementary science methods course supported preservice teachers in creating coherent inquiry science lessons around astronomy phenomena appropriate to elementary grades. We combined qualitative analysis of the preservice teachers' lesson plans and reflections with quantitative analysis of a pre–post astronomy content assessment to look for relationships between their ideas about inquiry, use of curricular resources, and relevant astronomy content knowledge with their success in developing coherent science inquiry investigations. This allowed us to better understand the preservice teachers' choices in lesson planning by converging on variables that influenced their design process and subsequently considering how the methods course influenced those variables.

Research Context: Elementary Science Methods Course

The study took place within three sections of an elementary science methods course at a small university in the northeastern United States. The course was primarily taught by

the first author, who was aided by a second methods instructor, not associated with this research. Preservice teachers' experiences were designed based on a social constructivist model of learning (Krajcik & Czerniak, 2007): key features included active engagement with phenomena, use and application of knowledge, engagement with multiple representations, participation in learning communities, and authentic tasks. The course was designed using a coherent science investigation framework with experiences sequenced to support preservice teachers in developing an understanding of this approach to science teaching. One level of coherence was achieved through a focused engagement in a single domain of science: the astronomy of celestial motion. During the first 5 weeks of the course, the preservice teachers were engaged in coherent science investigations around celestial motion phenomena. This investigation was followed by 5 weeks of fieldwork where the preservice teachers wrote and then taught lessons on the same celestial motion topics. We will discuss the nature of these two components of the course in more detail below. The course continued for five additional weeks; however, that portion of the course was not analyzed in the current study.

During the first 5 weeks of class, the preservice teachers were guided through an interrelated series of investigations in celestial motion, led by the driving question "How and why do celestial objects appear to move in the sky?" This theme was broken down into smaller investigation questions addressing the apparent motion of the Sun and stars and the changing lunar phases. The preservice teachers engaged in making first-hand observations of how shadows move, the Moon's appearance, and the stars at night. In class, the instructor led the preservice teachers in developing representations of how these objects move or appear to change over time. They also worked collaboratively to develop models that explain their observations, such as working with a globe and a lamp to explain the Sun and stars' apparent motion.

Their engagement in the methods course also included opportunities to critique an elementary astronomy curriculum, the *Sun, Moon, Stars* (Full Option Science System (FOSS), 2007). This practice of engaging preservice teachers in rounds of curriculum critiques has become an important feature of many reform-based science methods classes (e.g., Davis, 2006; Forbes, 2011; Gunckel, 2011). During these written critiques, preservice teachers were asked whether the FOSS curriculum would engage students in a coherent inquiry investigation and to suggest modifications for the lessons if they were to teach them during fieldwork. The instructor focused the preservice teachers' attention on how the curriculum and their own experiences conducting investigation in the course provided examples of how to engage children in scientific inquiry, with a particular focus on the connection between a scientific question, interrogating data for evidence, and constructing explanations to respond to the question and using evidence (NRC, 2000).

For the second 5-weeks of the course, preservice teachers wrote and taught a series of lessons, once a week for 5 weeks, in afterschool programs for K–6 grade students. Each lesson lasted approximately 45 minutes. One undergraduate section taught in an urban public school's afterschool program, whereas the other taught at a suburban private school; the graduate section taught at an urban environmental center. Each pair of preservice teachers worked with the same group of 4–12 children each week; children were grouped by grade level. The preservice teachers' assignment was to teach a single coherent science investigation during their time in the field; therefore, the pairs often worked together to develop an overall plan for the lessons. Each week, the preservice teachers had the following opportunities for feedback on their lessons: (1) written comments on the lesson plans from the second methods instructor, (2) whole-class discussion after each lesson was taught, and (3) written feedback by the primary methods instructor based on observation of their teaching. Finally, the preservice teachers wrote weekly reflections on the enactment of

the lessons. In their reflections, they discussed how they (or their partner) engaged their students in scientific inquiry during the lesson.

Participants

Every preservice teacher in each section of elementary science methods class participated in the study ($N = 30$). Participants were primarily female ($n = 29$) and European American ($n = 23$). We used pseudonyms for the participants, as listed in Table 2. Two sections were for undergraduates ($n = 18$), whereas the third was for graduate students in a postbaccalaureate certification program ($n = 12$); the instruction and assignments were identical across sections. The graduate students came from variety of backgrounds in terms of their bachelor degrees but were required to meet the same certification requirements as the undergraduates. One undergraduate participant had previous experience teaching preschool. Many of the preservice teachers in the graduate-level course had 1–2 years of experience teaching: six in preschool and one each in elementary and high school. Two participants remembered studying astronomy in high school, one had studied it in college, and two more indicated that they may have studied astronomy in high school. The rest indicated no prior study or only limited experience in elementary and/or middle school.

Data Sources

We collected and analyzed three forms of data for this study: five lesson plans per teaching pair, one pre/post astronomy content assessment per teacher, and five teaching reflections per teacher. The teachers were given the following instructions for their fieldwork assignment:

> Your goal is to develop and teach an inquiry investigation on elementary school astronomy. You and your partner will be writing lesson plans for each day of instruction. You and your partner should pick some aspect of the *FOSS Sun, Moon, and Stars* curriculum to focus on with the children. You may also use any other resources to help you design your instruction but choose materials that will help you develop an investigation around astronomy that relates to what we have been doing in our own investigations of astronomy.

Pairs cowrote the first of the five lessons; subsequent lessons were written and taught by one partner as the lead. While most teachers worked in pairs, one could not attend fieldwork at the same time as the others and did her fieldwork on a different day. This resulted in one trio of preservice teachers planning together in that section of the course. A total of 77 lesson plans were collected and analyzed.

After each lesson, each preservice teacher submitted a written reflection on their students' opportunities for participating in science inquiry (both for their own and their partner's lessons) in response to the following prompt:

> In your reflection, discuss where you saw students having the opportunity to participate in scientific inquiry as a part of your investigation. In other words, think back to what we have identified as "doing science as scientists do science" and discuss the ways you and your partner helped the students engage in doing science.

We analyzed a total of 148 reflections (two reflections were not submitted by one teacher).

An astronomy content assessment was given on the first day of class and again after fieldwork. The assessment combined items from two existing astronomy assessments: the *Astronomy Diagnostic Test* (Hufnagel, 2002) and *MOSART* (Sadler et al., 2010). Items were

selected to align to the specific conceptual areas targeted by the astronomy instruction in the first 5-weeks of the course, which were also the topics designated for the fieldwork lesson plans. Therefore, the assessment covered the apparent motion of the Sun, Moon, and stars; explanations for apparent celestial motion; change in lunar phases; explanations for lunar phases; and size and scale of the solar system and stars.

Data Analysis

We analyzed each pair's set of five lesson plans, examining the extent to which the preservice teachers were able to develop coherent science inquiry investigations and identifying the curriculum resources they used to develop their investigations. We then analyzed the individual reflections to help us understand the relationship between their understanding of scientific inquiry and their choices in planning coherent science inquiry investigations. Finally, we analyzed astronomy content assessments to investigate how preservice teachers' successes in developing coherent science inquiry investigations related to their relevant astronomy content knowledge. The detailed analysis procedures are provided below.

Lesson Plans. We developed a coding protocol that includes four categories relating engagement in inquiry-based investigations: investigation question, use of investigation question, data collection, and explanation. Within each of these categories, we developed codes that reflect *INSES,* as well as codes that arose from the nature of the choices the preservice teachers made in their lesson planning; these codes can be found in Supplemental Appendix A in the Supplementary Information. Some categories were applied to individual lessons, whereas others looked at connections *between* practices and *across* lessons. The authors compared coding for a subsets of 20 lesson plans until a minimum interrater reliability of Cohen's kappa = 0.8 was achieved for each code; however, one code continued to have a low Cohen's kappa after four rounds of comparison. The code "no connection" under the explanation category reached $\kappa = 0.64$. The authors both coded all lessons for "no connections," and all disagreements were subsequently resolved through discussion.

Codes were used to develop a rubric describing the levels of sophistication (LoS) of their science investigations. Research on teachers' design of inquiry often focuses on the level of openness in the investigation (Schwab, 1962). The level of openness is determined by the extent to which the student or the teacher determines each element of the investigation; more choices made by the students indicate a more open level of inquiry. Our LoS rubric did not focus on openness. Instead, our LoS rubric was based measuring the coherence of their investigations by examining the connections between features of inquiry (NRC, 2000): the extent to which the preservice teachers were able to design an investigation that posed a scientific question that led to constructing a scientific explanation based on evidence (Table 1).

As described in Table 1, investigations at LoS 1–3 start with a question or statement and lead students to engage with data in response to that question. At LoS 1 and 2, students are asked to engage in constructing explanations and, at LoS 1, the lesson plan describes how the teacher will guide students to make an explicit connection between evidence and explanation. LoS were developed through a combination of description of inquiry from INSES and trends observed in the students' lesson plans. For example, while LoS 1 describes a normative view of how a question leads to an explanation in scientific inquiry, we found that many lesson plans lacked an explicit connection between the data students collected and how it should be used as evidence in an explanation (LoS 2).

TABLE 1
Levels of Sophistication for Investigations Using *INSES* Criteria

Level	Description of Level	Lesson Plan Code[1]
1	Investigation question leads to examining data for evidence toward answering the question. An explanation is constructed in response to the investigation question and explicitly uses evidence as support.	All four codes must be connected to the same question/statement: **Investigation Question:** Question or Investigation statement **Use of IQ:** Data collection, responds to data **Data:** Connected **Explanation:** Connected
2	Investigation uses an investigation question that leads to interrogating data for evidence but leaves the connection between evidence and explanation *implied*.	All four codes must be connected to the same question/statement: **Investigation Question**: Question or Investigation statement **Use of IQ:** Data collection, responds to data **Data:** Connected Explanation: Implied
3	Lessons allow children to engage with data in response to a question, but children are not guided to construct an explanation in response to the investigation question.	All three codes must be connected to the same question/statement: **Investigation Question**: Question or Investigation statement **Use of IQ:** Data collection, responds to data **Data:** Connected
4	Children engage with data but not in response to a scientific question.	Does not include an investigation question or statement. **Data**: Not connected

[1]Detailed descriptions of categories and codes are given in Appendix A in the Supporting Information.

We also considered the number of lessons spent on coherent science investigations compared to the total number of lessons. To describe the extent to which they planned coherent science investigations across their five lessons, each preservice teacher was assigned a ratio score composed of the number of lesson plans spent on a LoS 1 coherent investigation (or investigations) divided by the total number of lesson plans. This is labeled as *coherence measure* in Table 2. For example, consider a pair with a LoS 1 investigation that spans three lesson plans plus two additional lesson plans with just hands-on activities. Their *coherence measure* would be three LoS 1 lessons divided by the total number of lessons (five), equaling 0.6.

Two additional codes described aspects of investigations that modified the sophistication of the lesson plans. *Verification lessons* were observed when the teacher stated the investigation question but followed this by answering the question (orally or with an activity) before engaging the students in collecting data and constructing an evidence-based explanation. We also identified lessons with *problematic content representations* that, if enacted, would lead students toward scientifically inaccurate conceptions or where the investigation methods described in the lesson could not be used to answer the investigation question posed.

TABLE 2
Inquiry Levels of Sophistication and Teaching Details for Each Preservice Teaching Pair

	Graduate (G) or Undergraduate (UG)	Coherence Measure[1]	Level 1	Level 2	Level 3	Level 4	Lesson Plan Topic(s)	Grade Level	Prior Teaching Experience
Carly[a] and Claire	UG	1	Five lessons				Explaining Sun's apparent motion	Fourth	None; none
Jackie[a] and Jade	G	1	Five lessons				Changing appearance of the Moon	K	1 year pre-K; 2.5 years pre-K
Luann and Lisa	G	1	Five lessons				Explaining phases of the Moon	Fourth	None; 1+ years pre-K
Owen and Olivia	G	1	Five lessons				Explaining Sun's apparent motion	Third to fourth	2+ years and MS; 1 year pre-K
Alisha and Adina	UG	0.8	Three lessons, one lesson, one lesson*				Explaining Sun's apparent motion (three lessons), Sun's appearance, seasons	Second	None; none
Eve and Erin	UG	0.6	One lesson, one lesson, one lesson				Seasonal constellations, circumpolar constellations, size and scale of solar system	Fourth to Fifth	None; none
Melanie[b]	G	0.6	One lesson				Phases of the Moon	Second	1 year pre-K; Kindergarten
Mica[b]		0.6	One lesson, one lesson						
Mandy[a,b]		0.2			One lesson	One lesson			None
Dana and Di	UG	0.4	Two lessons				Size and scale, Explaining Sun's apparent motion (two lessons), seasons, stars' motion	Third	None; none
Ivy and Isabel	UG	0.4	Two lessons		One lesson		Sun's apparent motion (two lessons), nature of planets, stars in day and night	K	4 years early learning center; none
Katherine and Kumaria	G	0.2	One lesson	One lesson, one lesson	Three lessons		Sun's path, day/night cycle, size and scale of Sun, Earth, Moon	K–1	5 years high school; Some pre-K

(*Continued*)

TABLE 2
Continued

	Graduate (G) or Undergraduate (UG)	Coherence Measure[1]	Level 1	Level 2	Level 3	Level 4	Lesson Plan Topic(s)	Grade Level	Prior Teaching Experience
Gina and Gwen	UG	0.2	One lesson		One lesson, one lesson		Size and scale, using ■ compass, apparent motion, constellations	K–1	None; none
Nadia	G	0		One lesson, three lessons**			Apparent motion of constellations, seasonal constellations (three lessons)	K–1	None
Flora and Fran	UG	0			One lesson, one lesson *	One lesson	Relative size of Earth and Moon, lunar phases, eclipses	Second to third	None; none
Heather and Hill	UG	0				Two lessons, one lesson, one lesson	Phases of the Moon (Two lessons), eclipses, seasonal and circumpolar stars	Sixth	None; none
Beth and Belinda	UG	0				One lesson	Day/night cycle, lunar phases	Third	None; none

Notes: * – *Scientifically inaccurate;* ** – *Verification lesson;* [a]Studied astronomy in high school or college. [b]Members of this trio had different coherence measures because the group wrote their first lesson together, followed by two lessons written by Melanie, then the group was split and the last two lessons were written by Mandy and Mica separately.
[1]Coherence measure: (Number of lessons at Level 1)/Total lessons.

Reflections. We used a grounded theory approach (Corbin & Strauss, 2007) to analyze preservice teachers' reflections. We began by broadly coding for instances of text (a sentence or a few sentences) that reflected their ideas about inquiry. To ensure rigor, we went through cycles of reviewing a sample of reflections, then discussing whether their ideas fit into existing codes or if new codes needed to be created. By the last rounds of reflection analysis, no new codes appeared suggesting that we had saturated our coding scheme. Three categories emerged from our open coding of these reflections: normative, alternative, and pedagogical ideas about inquiry (codes are included in Supplemental Appendix B in the Supplementary Information).

Normative ideas reflect how reform documents define engagement in scientific inquiry (NRC, 2000, 2012). Most of the codes in the *normative* category describe ways of engaging students in science practices, such as use of models in constructing explanations based on evidence, making predictions, or making a clear connection between question, evidence, and explanations. *Alternative* ideas are those not reflected in reform-based descriptions of inquiry or scientific practices. Some examples of *alternative* codes include the suggestion that children are doing inquiry when they engage in a hands-on activity, developing an explanation without the use of evidence, or making sure children get the right answer. Codes in the *pedagogical* ideas category describe methods or rational for designing learning environments for inquiry investigations. *Pedagogical* codes included the idea that the teacher should guide the students during their investigations and that student involvement in inquiry is enhanced when students have opportunities to ask their own investigation questions, personally collect data, and construct their own explanations. All of the *pedagogical* codes reflected normative views on teaching with an inquiry orientation (see Supplemental Appendix B in the Supporting Information). As with other studies of preservice elementary teachers (e.g., Biggers & Forbes, 2012), the teachers' reflections often focused on student-centered perspectives on inquiry-based elementary science.

After the reflections were coded, we compared each preservice teacher's ideas about inquiry with his or her inquiry LoS. We grouped teachers from highest to lowest, depending on the *coherence measure*, the number of lessons at LoS 1 divided by the total number of lessons: highest (four to five lessons of LoS 1 investigation, *coherence measure* = 0.8–1), upper middle (two to three lessons of LoS 1, 0.4–0.6), lower middle (one lesson of LoS 1, 0.2), and lowest (no LoS 1, 0). Then, we compared the number of reflections that they wrote about each of the three major categories of inquiry ideas (normative, alternative, and pedagogical) with their inquiry group. Finally, we examined the specific codes from each group's reflections to further illuminate the teachers' thinking about designing investigations.

Lesson Plan Resources. Each lesson plan included a list of curriculum resources the preservice teachers used to plan their lessons. We analyzed these lists by grouping the materials into four categories: the "FOSS *Sun, Moon, Stars* curriculum" that was assigned reading during the first part of the course, "other lesson plans" primarily found online, "class investigations" they participated in during their methods class, and other "conceptual resources" that provided content and context for their investigations. We used these categories to consider the extent to which the materials and resources provided by the methods course were used to inform their development of coherent science inquiry investigations.

Content Assessment. We calculated the internal consistency of items using Cronbach's alpha, leading us to drop three items from the assessment to increase the consistency of the measure. The resulting Cronbach's alpha reliability of the 21-item instrument was 0.702,

above the standard "rule of thumb" minimum of 0.7 (Brace, Kemp, & Snelgar, 2009). A paired *t* test was used to compare pre- and posttest results. To look for a correlation between the preservice teachers' relevant astronomy content knowledge and their development of coherent inquiry investigations, a Pearson's *r* test was used to compare pre- and posttest results with their *coherence measures*.

Limitations

The most salient limitations of this study concern our knowledge of the preservice teachers' prior science experiences, their enactment of lessons, and information on their future teaching plans. First, while we were able to generate a measure of their relevant astronomy knowledge, our analysis would have been improved with additional insight into their past science experiences to understand what may have influenced their ideas about science inquiry. Second, though prior research suggests teachers often implement the lessons they plan with a high degree of fidelity (Biggers, Forbes, & Zangori, 2013; Zangori, Forbes, & Biggers, 2013), we were not able to analyze their enactment of these lessons. Finally, we could improve our understanding of the participants' trajectories toward becoming well-started elementary teachers with additional data on their next cycle of planning coherent science inquiry investigations.

FINDINGS

In the following sections, we present findings from the lesson plan analysis illustrating the extent to which preservice teachers developed coherent science inquiry investigations. This is followed by analyses of three factors which prior research suggests may influence preservice teachers' use of inquiry in their lesson planning: the extent to which they drew on resources provided by their science methods course, their understanding of scientific inquiry as indicated by their weekly reflections, and their knowledge of relevant astronomy content as measured by the written pre/post assessments.

Coherent Science Inquiry Investigations: Lesson Plan Analyses

We analyzed each pair's set of five lesson plans to determine the extent to which they planned for coherent science inquiry investigations using the LoS rubric (Table 1). Table 2 shows the LoS for each pair's lessons and the number of lesson plans that each investigation extended. For example, Alisha and Adina wrote a three-lesson LoS 1 investigation and two single-lesson LoS 1 investigations (though one had problematic content representation). The table is organized, from top to bottom, in decreasing ratio of LoS 1 investigation lessons to total lessons (*coherence measure*). Table 2 also indicates whether the teacher was an undergraduate (UG) or graduate (G) student, the topics they chose for their lessons, the grade level of students, and their prior teaching experience.

To help communicate our findings, we will begin by describing the typical structure of a successfully developed LoS 1 inquiry investigation. The teachers began by posing *an investigation question* focused on a pattern of observations (such as the Sun's path or sequence of lunar phases). Children then made and recorded *observations* to determine this pattern, occasionally through direct observation of phenomena but often through use of a computer simulation or data sources brought in by the teacher. Teachers and students co-constructed an initial *explanation* that used the evidence to develop a *representation* of Earth-based observable patterns. These representations were often pictorial, but may also have been gestured or verbal descriptions. The investigation then continued as the

teacher encouraged children to think about *why* the observational pattern existed through a psychomotor and/or kinesthetic *modeling activity* (e.g., children pretending to rotate like the Earth to explain the Sun's daily path). This led to a new explanation that drew upon the representation developed earlier as evidence for designing or evaluating the space-based model of motion or observing orientation.

In some cases, the teachers' investigation ended after the development of the representation of the Earth-based observational pattern; such investigations, however, still would have counted as full LoS 1 investigations as the students were engaged in answering scientific questions with evidence-based claims. These LoS 1 investigations provided children with a coherent science inquiry experience around one science phenomenon.

First, we will consider how the teachers' lesson plans reflected coherent science inquiry experiences by planning LoS 1 investigations across the 5 weeks in their field placements. Four pairs of teachers created LoS 1 investigations that spanned all five lesson plans: They either spent all five lessons investigating the same investigation question or they started with an observational investigation question (such as *how* does the Sun appear to move?) and modified it during the lesson sequence to address a new aspect of the construct (*why* does the Sun appear to move?). One additional pair created a three-lesson LoS 1 investigation plus two more single-lesson LoS 1 investigations (though one of these had a problematic content representation). Six other pairs included LoS 1 investigations during one, two, or three of their lessons plans. Thus, more than two-thirds of the preservice teachers were able to write one or more lesson plans that engaged their students in a coherent scientific inquiry investigation and nearly half of the pairs spent three or more lessons on coherent investigations.

While the description of inquiry in LoS 1 is based on a normative goal of a coherent scientific inquiry investigation, the remaining LoS emerged as patterns in the preservice teachers' lessons. Pairs who created a LoS 2 investigation provided opportunities for students to construct explanations but did not make the use of evidence an explicit part of how children constructed explanations in the lesson plan. For example, teachers followed data collection and analysis with a broad summarizing question but no support or explicit guidance for selecting evidence to support the explanation. These were categorized as an implied use of evidence because of the temporal proximity to the data collection phase.

Other pairs only created LoS 3 or 4 investigations. LoS 3 investigations lacked an opportunity for children to construct an explanation using the data they collected in response to an investigation question. Beth and Belinda posed several potential investigation questions, but none led to collecting data. Instead, the students were engaged in a series of hands-on activities, such as making models out of play dough or coloring in worksheets to show the night side of Earth. Flora and Fran did not provide a question or clear statement to drive the investigation until their third and fourth lessons; when they did, the students did not have an opportunity to construct an explanation in response to the investigation question. Heather and Hill's LoS 4 investigations were limited by a lack of clear focus provided by a question or problem statement. Though the students engaged with potential data, there was no clear purpose to this work. The students were engaged with physical and kinesthetic models as well, but with the goal of reaching a "correct answer" through a hands-on activity.

We found additional problematic aspects within a few lessons. A few teachers wrote verification lessons that began with an appropriate investigation question, which was then answered by the teacher. Nadia created three lesson plans in which she answered the investigation prompt prior to allowing the students to observe and explain the phenomena for themselves. Two pairs included a lesson with problematic content representations. Flora wrote a lesson with an inaccurate explanation for lunar phases. Alisha planned a model to use in investigating the seasons that did not accurately represent this concept.

Table 2 also shows that not all lessons from each pair ended up on the LoS scale. Some teachers wrote lessons that focused on hands-on activities rather than engagement with any of the elements of coherent science inquiry investigations. Dana and Di wrote three such lessons. The first lesson engaged students in learning the relative size of the Earth and Moon through a directed modeling activity. Later lessons led the students in physical and kinesthetic models of the Earth–Sun system to learn why we have seasons and why we see different constellations at different times of the year. For example, the teachers proposed that students use flashlights to see how the length of day changes as the Earth orbits the Sun. Thus, the teachers engaged the students with a scientific model that could yield predictions which could be tested against real-world observations. However, the students did not have an opportunity to compare or develop the model in response to evidence. These lessons were promising in terms of engaging children in relevant content representations for astronomy but fell short of engagement in scientific inquiry due to the lack of connection to evidence.

Use of Curricular Resources

To explore potential influences on the preservice teachers' development of coherent science investigations and to consider why some preservice teachers were more successful than others, we considered the curriculum materials they drew upon in writing their lesson plans. Since the science methods course provided several potential sources for the preservice teachers to draw on, we considered the extent to which they drew on materials provided by the methods courses versus other external resources. There were three categories of resources that came from the science methods course: class investigations, FOSS curriculum, and conceptual resources. Nearly half of the preservice teachers (44%) cited specific investigations that they participated in as students during the first 5 weeks as the source of their lesson plans. However, few pairs cited the class investigations for more than one or two of their lessons.

Most pairs (63%) cited the use of the FOSS curriculum in one or more lessons, with five pairs from across the inquiry levels using the FOSS curriculum in three or more lessons (Cs, Os, Ks, Ns, and Bs). Our review of the FOSS curriculum found that, while the curriculum describes methods of engaging children in science practices, its support for coherent science inquiry investigations is limited. For example, while it does provide focus questions such as "How does the Sun move from sunrise to sunset?" the curriculum does not help the teacher see how these focus questions can be used to guide data collection and then be answered using evidence.

All pairs except Owen and Olivia (94%) included other significant *conceptual resources* in their lesson plans. These were resources that did not explicitly give teachers direct pedagogical ideas but may have shaped their lessons through the content and context they provided. For example, many pairs used the computer simulation *Stellarium* in their lessons. This resource provided an opportunity to simulate collecting data and thus was a likely a strong influence on the lessons they were able to design. Star charts and Moon charts were another frequently used resource and provided an opportunity for students to make comparisons of astronomical phenomena over time. Real photographs of the Sun and Moon were further resources gathered as data for investigations. Some groups also used reference materials, such as nonfiction texts on astronomy. All of these resources, with the exception of some nonfiction books used by Jackie and Jade, originated in investigations the teachers experienced in the methods course or in suggestions from the methods instructor.

Finally, most pairs (69%) also used *other lesson plans*—resources that they found on their own rather than encountered during the methods course. However, these were often used in only one or two lessons.

Returning to our question of how the curricular resources may have influenced the preservice teachers' development of coherent science inquiry investigations, we observed no patterns when comparing the category of curriculum materials cited and the fraction of lessons plans spent on coherent science inquiry investigations.

Understanding of Science Inquiry

We next used the analysis of preservice teachers' ideas about inquiry, as suggested by their weekly reflections, to further interpret the choices they made in writing coherent science inquiry investigation lesson plans. Figure 1 shows the number of days each preservice teacher's reflection was coded for each of the inquiry categories. Most preservice teachers (80%) spent more days identifying normative ideas about inquiry than alternative ideas and nearly a third (30%) did not include any alternative ideas in their reflections. Far fewer spent the same number of days on normative and alternative ideas (7%) or more days on alternative ideas over normative ideas (13%). Thus, our initial analysis suggests that most could recognize elements of scientific inquiry during science instruction when teaching or observing a peer teaching.

We investigated the relationship between their ideas about inquiry and the coherence of the investigations in their lesson plans. Figure 1 shows each preservice teacher sorted from highest to lowest *coherence measure*, and grouped as follows: *highest* group (four or five lessons at LoS 1, *coherence measure* of 0.8–1), *upper middle* group (two or three lessons at LoS 1, 0.4–0.6), *lower middle* group (one lesson at LoS 1, 0.2), and *lowest* group with no LoS 1 investigations across all lessons (*coherence measure* 0). Our analysis reveals that the *highest* group had only one pair of preservice teachers with a single instance of alternative ideas in their reflections, whereas the *lowest* group had frequent examples of alternative ideas on multiple days. The members of the *lowest* group each had more alternative instances than normative instances in their reflections. For example, Claire (*highest* group) had five reflections with normative ideas and had no instances coded as alternative or pedagogical ideas. In comparison Fran (*lowest* group) had two reflections with normative ideas, three reflections with alternative ideas, and three reflections with pedagogical ideas.

Another major difference between the *highest* and *lowest* groups is that the *highest* group had a greater diversity of normative ideas and these appeared across multiple days, whereas the *lowest* group's reflections included mostly "data collection" and "making predictions" codes, resulting in less variety of codes across fewer days. All teachers in the *highest* group had "constructing explanation based on evidence" or "modeling in constructing explanations based on evidence" in their normative beliefs in multiple reflections while few in the *lowest* group reflected on these sense-making practices. The *highest* group's reflections included evaluative practices suggesting a more in depth understanding of the practices of science, whereas the *lowest* group did not. These evaluative practices included engaging students in evaluating how useful and effective their data was for answering their research questions or evaluating whether or not their investigations had effectively answered their investigation questions. For example, Claire wrote: "next week, I plan to have the students revisit the investigation they created. I will have them analyze the results, and determine whether or not they felt that it was an effective investigation."

We compared the *highest* group to pairs that only developed single coherent inquiry investigation lesson plans (*lower middle*) to consider possible differences in their ideas about inquiry. The *highest* group had only one pair with a single reflection expressing an alternative belief while the *lower middle* group had pairs with multiple reflections including a variety of alternative beliefs (e.g., engaging in skills, hands-on engagement, and participating in a modeling activity without connection to evidence). In addition, we

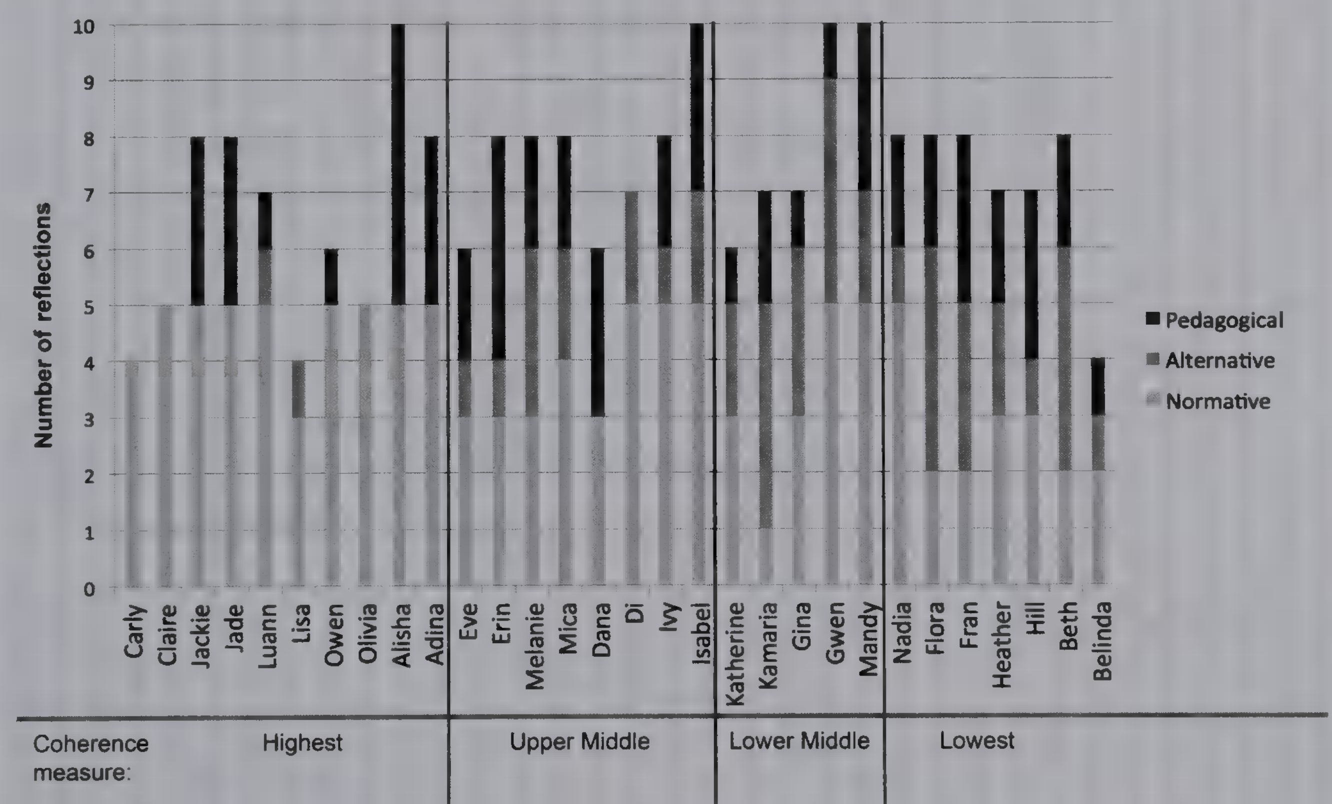


Figure 1. Preservice teachers are grouped by the *coherence measure* calculated with the number of lessons in LoS 1 divided by their total number of lesson plans (highest is *coherence measure* = 0.8–1, upper middle = 0.4–0.6, lower middle = 0.2, and lowest = 0) along the *x* axis. The *y* axis shows the number of reflections in which each preservice teacher wrote about each of the three inquiry idea categories: normative, alternative, and pedagogical.

found that while the preservice teachers in the *lower middle* group reflected on sense-making practices, such as constructing explanations based on evidence, this occurred less frequently than in the *highest* group.

We also considered the extent to which the preservice teachers reflected on the connected nature of science practices as a possible explanation for why some preservice teachers developed coherent science inquiry investigations while others did not. Carly and Claire reflected on the connection between question, evidence, and explanation across multiple reflections. Both reflected on their investigation question after the first lesson and indicated their plan for a 5-day coherent inquiry investigation. Carly wrote,

> we decided our driving question will be "How and why do the sun, moon, and stars move?" This will also lead us right to scale.... I first facilitated a discussion with the students about what we have learned so far from all the data we have gathered on the topic and then I let them get to work on modeling what they have learned from the data. (Reflection 1)

In this excerpt, she reflected on the inquiry question, data they gathered, and explanation through modeling based on evidence:

> I believe that this practice in modeling helped improve their inquiry skills, particularly the importance of backing up their points with reasoning and evidence because by the time of the presentations, students were asking the other groups questions like, 'How does that show the sun rises in the east?' or 'What did you mean by when you said shift?' They have begun to expect that students have evidence to support their claims. I was also glad that the students were able to show how the earth rotates. (Carly, Reflection 4)

However, only Carly and Claire consistently reflected in a way that connected their investigation question, evidence, and explanation development. And while others reflected on similar connections in one or two reflections, most of the preservice teachers' reflections did not explicitly discuss these connections.

Relevant Astronomy Content Knowledge

Finally, we examined the preservice teachers' astronomy knowledge as a potential explanation for why some pairs developed coherent science inquiry investigations and others did not. Initially, the preservice teachers had a relatively low level of knowledge of elementary astronomy, with a mean score of 7.7 ($SD = 2.7$) out of 21 on the content assessment. The mean score on the posttest, administered 3 weeks after fieldwork, was 12.3 ($SD = 3.3$), a statistically significant improvement ($t = 9.97$, $p < .001$). While we observed no correlation between their initial knowledge of astronomy and their *coherence measure*, there was a significant correlation between their postinstruction astronomy knowledge and their *coherence measure* (Pearson's $r = .366$, $p < .05$). This suggests that what they learned during the initial 5 weeks of investigation and during the process of developing lessons was related to their construction of coherent science inquiry investigations.

DISCUSSION

Our study examined preservice elementary teachers' development of coherent science inquiry investigations and factors that may help explain their success. We developed and applied a method of coding preservice teachers' lesson plans to evaluate whether they had planned a coherent science inquiry lesson within and across multiple lessons. Our

findings suggest that, with appropriate support provided by a science methods course, many preservice elementary teachers are able to develop coherent science inquiry investigations designed to engage students in constructing explanations about astronomical phenomena. A third of the pairs were able to develop coherent investigations for all five lessons they taught in an afterschool program, starting with an investigation question that led to collecting and analyzing data and co-constructing explanations based on evidence. These preservice teachers' development of coherent science inquiry investigations suggests they are beginning to consider methods that engage students deeply with a single concept over time rather than a more fragmented approach of hands-on "science activities that work" (Appleton, 2002, 2003).

However, not all pairs developed coherent science investigations for all or most of their lessons. One third of the preservice teachers developed either a single lesson plan with a coherent science inquiry investigation (LoS 1) or no lessons at that level of coherence. Even when preservice teachers shift toward more inquiry-oriented approaches, moving toward belief in the centrality of evidence-based explanations in science is still difficult (Zembal-Saul, 2009). In other ways, these results are promising, considering the fact that these were the preservice teachers' first experiences in attempting to develop coherent inquiry investigations. Even the students who did not develop LoS 1 investigations engaged their students in opportunities to collect and discuss data as well as other scientific practices. Thus, we see an indication of a potential continuum between preservice teachers who were partially successful (including some elements of scientific inquiry, such as opportunities for collecting data, but primarily focused on hands-on activities) to those who wrote fully developed, coherent investigations across all of their lesson plans.

Preservice teachers need to experience the process of learning science in the ways we hope they will teach science in their future classrooms (Haefner & Zembal-Saul, 2004; Zembal-Saul, 2009). Thus, the science methods course was designed to support the preservice teachers' development of coherent science investigations through experiences with coherent science inquiry investigations and by providing examples of curricula and other resources they could use to develop inquiry-based astronomy lesson plans. These experiences helped them to both understand the features of inquiry and develop the type of coherent understanding of astronomy needed to teach elementary students. We will discuss how our analysis suggests preservice teachers utilized these experiences and resources to inform their choices in planning science lessons.

Preservice Teachers' Use of Resources and Curriculum

Prior research suggests that experience engaging in scientific inquiry is an important predictor of whether new teachers will implement inquiry-based investigations in their teaching (Windschitl, 2002). Furthermore, long-term engagement in authentic science investigations can also shift preservice teachers' thinking toward recognizing the centrality of questions to science (Haefner & Zembal-Saul, 2004). Therefore, we considered the role the methods course played in supporting preservice teachers in planning for coherent science inquiry investigations through their engagement in coherent science investigations during the first 5 weeks of the course. Close to half of the pairs indicated they were drawing on the investigations from the first 5 weeks of class as a source of one or more of their lesson plans. Furthermore, many preservice teachers used resources provided by the methods course, such as the computer simulation *Stellarium* and access to monthly star charts.

Many of the preservice teachers also cited the FOSS *Sun, Moon, and Stars* curriculum, chosen for the class to read and critique because it covered the relevant astronomy content and has an inquiry-based approach. Forbes and Davis (2010a) suggest that the most

influential factor in determining the level of inquiry of preservice elementary teachers' lessons is the curriculum material they choose to adapt. Thus, we considered how this particular FOSS curriculum may have supported a coherent science inquiry approach to lesson planning. Our analysis suggests that, despite having an inquiry focus, it also has limitations that would not have supported teachers in developing coherent investigations, according to our criteria, without significant adaptation from the teachers. In particular, the curriculum did not include clearly stated investigation question leading to collecting data and constructing evidence-based explanations. Research on elementary teachers suggests that their curriculum implementations often do not go beyond the level of inquiry in the original lessons (Biggers et al., 2013; Zangori et al., 2013); those preservice teachers who relied on the FOSS curriculum may have needed additional support to write coherent science inquiry lesson plans.

Preservice Teachers' Ideas About Inquiry

Teachers' ideas about inquiry are reflected in the ways they enact science in the classroom (e.g., Furtak & Alonzo, 2010). The preservice teachers' ideas about what counts as "doing scientific inquiry" were related to the number of lesson plans they wrote which engaged children in coherent science inquiry investigations. Preservice teachers who developed these coherent investigations across many lesson plans identified normative scientific practices in their reflection and discussed more sense-making practices than other preservice teachers. Those who did not develop coherent science inquiry investigations across most of their lesson plans often indicated ideas that do not correspond to normative descriptions of inquiry; these alternative ideas have been observed in other studies of preservice teachers, such as thinking that hands-on activities are sufficient to count as inquiry investigations (Biggers & Forbes, 2012; Davis & Smithey, 2009; Haefner & Zembal-Saul, 2004). Roehrig and Luft (2004) described secondary science teachers as ranging from inquiry teachers, to process-oriented teachers, to traditional teachers. We observed a focus on process orientation, in which instructional activities are designed to help students learn science skills, in some of the alternative ideas of our preservice teachers though none appeared to be writing lessons from a purely traditional perspective. However, even when the teachers in our study revealed alternative views of engaging students in inquiry, many of their lesson plans and ideas about inquiry focused on student-centered instruction. Such conceptualizations are a crucial element for early career teachers to eventually develop inquiry-based classrooms (Roehrig & Luft, 2004).

The literature on teachers' ideas about scientific inquiry can be organized around two dimensions (Furtak, Seidel, Iverson, & Briggs, 2012): a *conceptual dimension,* which includes the conceptual structures, epistemic frameworks, social interactions, and procedural methods of inquiry, and a *guidance dimension*, which examines the continuum between teacher led, teacher-guided, student-led, or discovery approaches. While we anticipated capturing preservice teachers' views on the *conceptual dimension* in their reflections, we also found that the teachers chose to reflect on the *guidance dimension*—what we referred to as "pedagogical ideas." This focus on the pedagogical aspects of inquiry is not surprising given their experience in the methods course. During the methods course, the students engaged in scientific inquiry investigations and were guided to think about the epistemological features of their experiences. These investigations were also blended with instruction that emphasized the pedagogical benefits of engaging children in inquiry-based instruction through a socioconstructivist framework (Krajcik & Czerniak, 2007). Our findings are consistent with previous research suggesting preservice teachers "typically describe inquiry as important to incorporate to promote student interest, not to engage students in genuine

scientific activity" (Davis, 2006, p. 348). This suggests that many of the preservice teachers were highly attuned to factors that will help their students learn but not necessarily for the sake of engaging in a coherent science inquiry investigation. However, our analysis does not suggest these views either hindered or supported their development of coherent inquiry investigations, as these pedagogically focused reflections were prevalent across all the preservice teachers.

Preservice Teachers' Relevant Astronomy Knowledge

Teachers need a deep and flexible knowledge of the content to plan effective instruction (Borko, 2004). The teachers significantly improved their astronomy content knowledge after their experiences in the course. We also found a positive correlation between preservice teachers' postinstruction content knowledge and their *coherence measure*, thus adding to the literature showing that teachers need sufficient science content knowledge to develop inquiry-based instruction (e.g., Roehrig & Luft, 2004). This correlation suggests that the coherence and sequencing of the science content focus (i.e., astronomy), through the investigations during the first 5 weeks of the methods course and their lesson plan writing during the second 5 weeks, may have contributed to their success.

A significant body of research suggests that both children (e.g., Plummer, Kocareli, & Slagle, 2014) and adults (e.g., Plummer, Zahm, & Rice, 2010; Zeilik & Bisard, 2000; Zeilik & Morris, 2003) find astronomy challenging to learn, including elementary content standards. Few teachers in our study studied astronomy in high school or college, and many did not ever remember studying astronomy; this is not surprising as astronomy may be left out of the curriculum in many middle and high schools (Plummer & Zahm, 2010). This trend suggests that without significant support in developing relevant content knowledge, elementary teachers are unlikely to engage their students in coherent science inquiry investigations.

CONCLUSION

The *NGSS* emphasize engaging students in a breadth of science practices across core disciplinary ideas and across grade levels. This goal is undeniably important, but to fully develop students' understanding of the scientific enterprise, teachers should also sequence students' experiences with science practices in coherent investigations. This study goes beyond previous studies by explicitly investigating the potential and challenges associated with preservice elementary teachers developing coherent science inquiry investigations. We explicitly examined the connections they made between scientific questions, data collected, and explanation construction in their lesson plans. By analyzing their writing across five consecutive lesson plans, we were able to analyze their level of coherence, whether it took place within a single lesson or across multiple lessons. Our findings suggest that many preservice teachers are capable of developing these coherent science inquiry investigations, given support for their understanding of science inquiry practices and knowledge of relevant science content.

In this study, we considered the ways in which the design of the science methods course provided this support. One critical feature of the course was the coherence and sequencing of the preservice teachers' experiences during the first 5 weeks of the course and the second 5 weeks where they engaged in planning and teaching lessons to elementary students. Part of this coherence was the focus on a single science content area, observational astronomy; the preservice teachers participated in astronomy investigations followed by opportunities to apply what they learned by writing and teaching astronomy lessons to elementary students.

The differences in how scientists conduct investigations in different domains may make it difficult to compare teachers focusing on different science disciplines (Roth et al., 2006); for example, while in some fields, scientists conduct carefully controlled experiments to test their hypotheses, other fields, such as astronomy, are based on descriptive methodologies. When school science focuses on the "scientific method," students are often only exposed to experimental studies where one variable is tested in comparison to a control group, a methodology not possible in many science fields (Windschilt, Thompson, & Braaten, 2008). Our focus on observational astronomy across the preservice teachers' experiences allowed us to reduce potential variation due to differences between doing and teaching science across different domains of science within the study.

For those preservice teachers who did develop coherent science inquiry investigations in their lesson plans, the science methods course appears to have influenced them in two primary ways. First, teachers with higher astronomy content knowledge at the end of their lesson planning experiences were more likely to produce coherent science inquiry investigations in their lesson plans. Though the literature often points out the importance of content knowledge for successful teaching, few studies have examined correlations between teachers' use of inquiry and their content knowledge (Luera et al., 2005). By narrowly focusing on one area of science across all the teachers in this study, we were able to provide evidence of this important connection.

Second, teachers whose reflections included fewer alternative ideas about inquiry and more descriptions of sense-making practices spent more of their lesson plans on coherent science inquiry investigations than other teachers. Thus, preservice teachers' reflections on inquiry further reveal their ideas of what counts as scientific inquiry and point to important aspects of their decision-making processes when planning science lessons. Both of these correlations led us to emphasize the importance of providing coherence across and connection between the preservice teachers' experiences in the methods course and their fieldwork opportunities.

Implications

The preservice teachers' success in planning coherent science inquiry investigations points to several implications for the design of elementary science methods courses. First, we recommend that elementary science methods courses support preservice teachers by engaging them in inquiry-based science curricula as students, followed by the opportunity to use that experience to plan and teach science lessons to elementary students (Haefner & Zembal-Saul, 2009). By focusing these experiences in the same science content area, the preservice teachers will deepen their understanding of relevant science content knowledge and be better prepared to translate their own science learning experiences into coherent investigations ready for elementary students during fieldwork. This conceptual coherence between methods course and fieldwork assignment provides opportunities for the methods instructor to support preservice teachers through in-class investigation examples, opportunities to critique relevant curricula, and opportunities to develop a community of teachers working toward closely aligned goals.

Second, we recommend that methods courses plan opportunities for preservice teachers to reflect on and identify their own use of inquiry in the lessons they write and teach. In addition to the importance of reflective practices to teachers' growth (Abell & Bryan, 1997; Bryan & Abell, 1999; Crawford, 1999; Davis et al., 2006; Lotter et al., 2009; Singer, 2005; Zembal-Saul et al., 2000), these reflections also provide a window into the preservice teachers' understanding of scientific inquiry beyond what is revealed in their lesson plans. Methods instructors may identify alternative ideas in these reflections and

plan for opportunities to address these problematic conceptualizations in future classes. We further recommend that instructors direct their preservice teachers to reflect on the connections between individual lessons to emphasize the importance of building a coherent science curriculum over time rather than focusing on individual lessons (e.g., NRC, 2012; Roth & Garnier, 2006). This was not explicitly asked of our preservice teachers, and few reflected on the connections between their lessons.

Third, preservice teachers' experiences, both engaging in investigations as learners and planning as teachers, should emphasize the importance of sense-making practices involved in developing evidence-based explanations. Supporting preservice teachers in developing evidence-based explanations and modeling practices through in-class investigations will better prepare them to attempt these practices in their fieldwork experiences. However, we note that many of the preservice teachers in our study either missed opportunities to include these sense-making practices in their lessons or to discuss this aspect of inquiry in their reflections. Thus we suggest making reflection on sense-making practices an explicit part of preservice teachers' reflections on their fieldwork to help them focus on this vital aspect of doing science. If fieldwork teaching experiences are spaced out over time, such as once-a-week lessons in this study, the methods instructor would have the opportunity to address preservice teachers' ideas about sense-making practices as exhibited in their lessons and reflections.

Finally, given the influence that school culture can play on new teachers' pedagogical choices (Avraamidou & Zembal-Saul, 2010; Forbes, 2013; McGinnis, Parker, & Graeber, 2004), we suggest that using nonformal settings for fieldwork allows preservice teachers more flexibility in their lesson planning. In this study, the preservice teachers designed for and taught in afterschool programs. This out-of-school setting allowed the preservice teachers to focus on elements of teaching without the additional pressures of traditional school placements. For example, the preservice teachers may have been more successful in developing the type of coherent inquiry investigations proposed by their methods course instructor because they were not receiving conflicting requirements from a classroom mentor teacher. Given the influence of early field experiences on preservice teachers' dispositions toward teaching, it is essential that these experiences be supportive of inquiry-based practices (Forbes, 2013).

Our method of analyzing preservice teachers' use of inquiry also has implications for teacher education and research. While previous studies have considered the extent to which teachers include features of scientific inquiry in their lessons and teaching (e.g., Biggers & Forbes, 2012; Forbes, 2011, 2013; Forbes & Davis, 2010a), we focused primarily on the connections the teachers made between features of inquiry as central to our analysis. Our LoS rubric (Table 1) could be used as an instructional tool for methods professors to guide preservice teachers' thinking about planning investigations, as it is broad enough to include multiple investigative methods. The rubric could also be used as an analytical tool for researchers in studying how teachers develop and implement coherent science inquiry investigations across age groups and across scientific disciplines.

These findings have further implications for the development of educative curricula, designed to support both teacher and student learning (Davis & Krajcik, 2005; Schneider & Krajcik, 2002). Educative curricula should highlight the deeply connected nature of the experiences in a scientific investigation to move away from a piecemeal perspective reflected in hands-on and skill-based lessons. Such curricula could build on the concept of the *CSCS* (Roth & Garnier, 2006), by helping the teacher learn methods of engaging children in science practices toward a deepening understanding of a single conceptual goal. In particular, this type of curricula may have helped some of the preservice teachers in this study who had limited astronomy knowledge. A lack of understanding of the connections

between concepts in astronomy may have limited their ability to envision the steps of an investigation that builds toward an evidence-based explanation. In doing so, educative curricula should provide explicit examples of the *types of investigation questions* that are feasible for elementary astronomy investigations and *how to connect* those questions to appropriate data collection methods. Lesson plans that were not classified in the more sophisticated levels of inquiry-based investigation often did not connect a clear investigation question with engagement in data collection. Slater, Slater, and Shaner (2008) suggest that generating scientific questions that are both ready and worthy to investigate is the most difficult element of the scientific process for students. However, science curricula often do not provide teachers with questions that support scientific investigations (Forbes & Davis, 2010b; Kesidou & Roseman, 2002). Another limitation we observed in some preservice teachers' lessons was the lack of explicit connection between their plans for their students to collect data and opportunities to construct explanations. Thus, curricula should provide explicit support for teachers in how to support students in moving from collecting data to using evidence purposefully in constructing an explanation.

The authors wish to thank Carla Zembal-Saul and Jerry Weiner for their assistance on this project. An earlier version of this work was presented at the 2011 conference of the Astronomical Society of the Pacific and the 2013 international conference of the National Association for Research in Science Teaching.

REFERENCES

Abell, S. K., & Bryan, L. A. (1997). Reconceptualizing the elementary science methods course using a reflection orientation. Journal of Science Teacher Education, 8, 153–166.

Akerson, V. L., Morrison, J. A., & McDuffie, A. R. (2006). One course is not enough: Preservice elementary teachers' retention of improved views of nature of science. Journal of Research in Science Teaching, 43, 194–213.

Alonzo, A. (2002). Evaluation of a model for supporting the development of elementary school teachers' science content knowledge. Presented at Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science, Charlotte, NC.

Appleton, K. (2002). Science activities that work: Perceptions of primary school teachers. Research in Science Education, 32(3), 393–410.

Appleton, K. (2003). How do beginning primary school teachers cope with science? Toward an understanding of science teaching practice. Research in Science Education, 33(1), 1–25.

Avraamidou, L., & Zembal-Saul, C. (2010). In search of well-started beginning science teachers: Insights from two first-year elementary teachers. Journal of Research in Science Teaching, 47, 661–696.

Biggers, M., & Forbes, C. (2012). Balancing teacher and student roles in elementary classrooms: Preservice elementary teachers' learning about the inquiry continuum. International Journal of Science Education, 34, 2205–2229.

Biggers, M., Forbes, C. T., & Zangori, L. (2013). Elementary teachers' curriculum design and pedagogical reasoning for supporting students' comparison and evaluation of evidence-based explanations. The Elementary School Journal, 114, 48–72.

Brickhouse, N. W. (1990). Teachers' beliefs about the nature of science and their relationships to classroom practice. Journal of Teacher Education, 41(3), 53–62.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. Educational Researcher, 33, 3–15.

Brace, N., Kemp, R., & Snelgar, R. (2009). SPSS for psychologists. Basingstoke, England: Palgrave Macmillan.

Bryan, L. A., & Abell, S. K. (1999). Development of professional knowledge in learning to teach elementary science. Journal of Research in Science Teaching, 36(2), 121–139.

Corbin, J., & Strauss, A. (2007). Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.). Thousand Oaks, CA: Sage.

Crawford, B. A. (1999). Is it realistic to expect a preservice teacher to create an inquiry-based classroom? Journal of Science Teacher Education, 10(3), 175–194.

Davis, E. A. (2006). Preservice elementary teachers critique of instructional material for science. Science Education, 90, 348–375.
Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. Educational Researcher, 34(3), 3–14.
Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. Review of Educational Research, 76(4), 607–651.
Davis, E. A., & Smithey, J. (2009). Beginning teachers moving toward effective elementary science teaching. Science Education, 93, 745–770.
Forbes, C. (2011). Preservice elementary teachers' adaptation of science curriculum materials. Science Education, 95, 1–29.
Forbes, C. (2013). Curriculum-dependent and curriculum-independent factors in preservice elementary teachers' adaptation of science curriculum materials for inquiry-based science. Journal of Science Teacher Education, 24, 179–197.
Forbes, C., & Davis, E. (2010a). Curriculum design for inquiry: Preservice elementary teachers' mobilization and adaptation of science curriculum materials. Journal of Research in Science Teaching, 47(7), 365–387.
Forbes, C. T., & Davis, E. A. (2010b). Beginning elementary teachers' beliefs about the use of anchoring questions in science: A longitudinal study. Science Education, 94(2), 365–387.
Full Option Science System (2007). Sun, moon, and stars. Berkeley, CA: The Lawrence Hall of Science.
Furtak, E. M., & Alonzo, A. C. (2010). The role of content in inquiry-based elementary science lessons: An analysis of teacher beliefs and enactment. Research in Science Education, 40(3), 425–449.
Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching a meta-analysis. Review of Educational Research, 82(3), 300–329.
Gess-Newsome, J. (1999). Teachers' knowledge and beliefs about subject matter and its impact on instruction. In J. Gess-Newsome & N. G. Lederman (Eds.), Examining pedagogical content knowledge: The construct and its implications for science education (pp. 51–94). Dordrecht, The Netherlands: Kluwer.
Gunckel, K. L. (2011). Mediators of a preservice teacher's use of the inquiry-application instructional model. Journal of Science Teacher Education, 22(1), 79–100.
Haefner, L., & Zembal-Saul, C. (2004). Learning by doing? Prospective elementary teachers' developing understandings of scientific inquiry and science teaching and learning. International Journal of Science Education, 26(13), 1653–1674.
Hapgood, S., Magnusson, S., & Palincsar, A. (2004). Teacher, text, and experience: A case of young children's scientific inquiry? Journal of the Learning Sciences, 13, 455–505.
Hufnagel, B. (2002). Development of the astronomy diagnostic test. Astronomy Education Review, 1(1), 47–51.
Jones, M. G., & Edmunds, J. (2006). Models of elementary science instruction: Roles of science specialists. In Appleton, K. (Ed), Elementary Science Teacher Education: International Perspectives on Contemporary Issues and Practice (pp. 317–343). Mahwah, NJ: Erlbaum.
Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. Journal of Research in Science Teaching, 39(6), 522–549.
Krajcik, J., & Czerniak, C. (2007). Teaching science in elementary and middle school: A project-based approach. New York: Routledge.
Lotter, C., Singer, J., & Godley, J. (2009). The influence of repeated teaching and reflection on preservice teachers' views of inquiry and nature of science. Journal of Science Teacher Education, 20, 553–582.
Luera, G. R., Moyer, R. H., & Everett, S. A. (2005). What type and level of science content knowledge of elementary education students affect their ability to construct an inquiry-based science lesson?. Journal of Elementary Science Education, 17(1), 12–25.
McGinnis, R., Parker, C., & Graeber, A. O. (2004). A cultural perspective of the induction of five reform-minded beginning mathematics and science teachers. Journal of Research in Science Teaching, 41(7), 720–747.
Melville, W., Fazio, X., Bartley, A., & Jones, D. (2008). Experience and reflection: Preservice science teachers' capacity for teaching inquiry. Journal of Science Teacher Education, 19(5), 477–494.
Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. Cognition and Instruction, 22(2), 219–290.
National Research Council. (2000). Inquiry and the National Science Education Standards. Washington, DC: National Academies Press.
National Research Council. (2007). Taking science to school: Learning and teaching science in Grades K–8. Washington, DC: The National Academic Press.
National Research Council (2012). A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. Washington, DC: National Academies Press.
NGSS Lead States. (2013). The Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.

Plummer, J. D., Kocareli, A., & Slagle, C. (2014). Learning to explain astronomy across moving frames of reference: Exploring the role of classroom and planetarium-based instructional contexts. International Journal of Science Education, 36, 1083–1106.

Plummer, J. D., & Zahm, V. (2010). Covering the Standards: Astronomy teachers' preparation and beliefs. Astronomy Education Review, 9(1), 010110.

Plummer, J. D., Zahm, V., & Rice, R. (2010). Inquiry and astronomy: Preservice teachers' investigations in celestial motion. Journal of Science Teacher Education, 21, 471–493.

Roehrig, G. H., & Luft, J. A. (2004). Constraints experienced by beginning secondary science teachers in implementing scientific inquiry lessons. International Journal of Science Education, 26, 3–24.

Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., et al. (2006). Teaching science in five countries: Results from the TIMSS 1999 video study (NCES 2006-011). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Roth, K., & Garnier, H. (2006). What science teaching looks like: An international perspective. Educational Leadership, 64(4), 16.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. Journal of Research in Science Teaching, 48(2), 117–148.

Sadler, P., Coyle, H., Miller, J., Cook-Smith, N., Dussault, M., & Gould, R. (2010). The astronomy and space science concept inventory: Development and validation of assessment instruments aligned with the K–12 national science standards. Astronomy Education Review, 8(1), 010111.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). A splintered vision: An investigation of U.S. science and mathematics education. Dordrecht, The Netherlands: Kluwer.

Schneider, R. M., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. Journal of Science Teacher Education, 13(3), 221–245.

Schneider, R., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. Review of Educational Research, 81, 530–565.

Schwab, J. J. (1962). The teaching of science as inquiry. In J. J. Schwab & P. F. Brandwein (Eds.), The teaching of science (pp. 3–103). Cambridge, MA: Harvard University Press.

Singer, J. (2005). Integrating technology and pedagogy: The ideas, the shift and the targets. In S. Rhine & M. Bailey (Eds.), Integrated technologies, innovative learning: Insights from the PT3 Program (pp. 199–215). Eugene, OR: International Society for Technology in Education.

Siry, C., Ziegler, G., & Max, C. (2012). "Doing science" through discourse-in-interaction: Young children's science investigations at the early childhood level. Science Education, 96, 311–366.

Slater, S. J., Slater, T. F., & Shaner, A. (2008). Impact of backwards faded scaffolding in an astronomy course for pre-service elementary teachers based on inquiry. Journal of Geoscience Education, 56(5), 408–416.

Varelas, M., Pappas, C., Kane, J., Arsenault, A., Hanks, J., & Cowan, B. (2008). Urban primary-grade children think and talk science: Curricular and instructional practices that nurture participation and argumentation. Science Education, 92, 65–95.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. Review of Educational Research, 72, 131–175.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. Science Education, 92, 941–967.

Zangori, L., Forbes, C. T., & Biggers, M. (2013), Fostering student sense making in elementary science learning environments: Elementary teachers' use of science curriculum materials to promote explanation construction. Journal of Research in Science Teaching, 5, 989–1017.

Zeilik, M., & Bisard, W. (2000). Conceptual change in introductory-level astronomy courses. Journal of College Science Teaching, 29, 229–232.

Zeilik, M., & Morris, V. J. (2003) An examination of misconceptions in an astronomy course for science, mathematics, and engineering majors. Astronomy Education Review, 2(1), 101–119.

Zembal-Saul, C. (2009). Learning to teach elementary school science as argument. Science Education, 93, 687–719.

Zembal-Saul, C., Blumenfeld, P., & Krajcik, J. (2000). Influence of guided cycles of planning, teaching, and reflection. Journal of Research in Science Teaching, 37, 318–339.

Zembal-Saul, C., McNeill, K. L., & Hershberger, K. (2012). What's your evidence? Engaging K–5 students in constructing explanations in science. Boston: Pearson.

Science
Education

LEARNING

Promoting Conceptual Coherence Within Context-Based Biology Education

MICHA H. J. UMMELS,[1] MARCEL J. A. KAMP,[1] HANS DE KROON,[2]
KERST TH. BOERSMA[3]
[1]*Radboud Graduate School of Education, Radboud University, 6525 HP Nijmegen, The Netherlands;* [2]*Institute for Water and Wetland Research, Radboud University, 6525 HP Nijmegen, The Netherlands;* [3]*Freudenthal Institute for Science and Mathematics Education, University of Utrecht, 3584 CC Utrecht, The Netherlands*

Received 11 March 2014; accepted 14 April 2015
DOI 10.1002/sce.21179
Published online 14 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: In secondary science education, the learning and teaching of coherent conceptual understanding are often problematic. Context-based education has been proposed as a partial solution to this problem. This study aims to gain insight into the development of conceptual coherence and how context-embedded learning-teaching activities (LT) can promote this. We describe a case study in which a context-based lesson sequence about protein-rich food production was designed and conducted in a 10th-grade biology class. The conceptual framework consisted of transformations of forms of energy and matter in photosynthesis, cellular respiration, and biosynthesis. All relevant concepts and their interconnections (propositions) were captured in a reference concept map. A research scenario was used to evaluate whether the lesson sequence was conducted as intended. Learning outcomes were determined by analyzing written products on the occurrence of propositions from the reference concept map. Additional interviews provided insight into the development of conceptual coherence in relation to three context-embedded LT activities: using graphic visualizations, writing, and concept mapping. The results indicated that students improved in mentioning propositions from the reference concept map. Propositions relating metabolic processes and including forms of energy were still difficult. Finally, successful elements of the three LT activities are considered. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**958–985, 2015

Correspondence to: M. H. J. Ummels; e-mail: m.ummels@docentenacademie.ru.nl

INTRODUCTION

Research on learning and teaching natural science has shown that students' conceptual knowledge at all educational levels is often incoherent (e.g., DiSessa, Gillespie, & Esterly, 2004; Wandersee, Mintzes, & Novak, 1994). This lack of coherence is reflected by students' inability and inconsistency to retrieve and connect concepts: for example, students often have difficulty explaining and predicting natural phenomena and events. Moreover, students have difficulties transferring concepts to other situations than those in which they were learned (Bransford, Brown, & Cocking, 2000b). Because traditional teaching and learning approaches are often inappropriate in terms of helping students to assimilate coherent frameworks of concepts, an international trend toward context-based education has developed in science education (e.g., Gilbert, 2006).

Context-based approaches generally aim to improve students' engagement by situating science learning in real-world contexts (King & Ritchie, 2012). This framing helps students to appreciate the role science plays in their own lives and in society. Because various concepts come together within a context and reappear in other contexts, they are assumed to provide a basis for the development of coherent mental maps of the relationships between them (Gilbert, 2006). In this paper, we refer to the term "conceptual coherence" as the ability of a person's cognitive network to establish meaningful connections between concepts.

Until now, there has been limited empirical evidence proving that context-based education has a significant impact on the development of students' conceptual coherence (Bennett, Lubben, & Hogarth, 2007). Tsai (2000) found that a science–technology–society instructional approach, similar to a context-based approach, improved the extent, richness, and connectivity of students' cognitive structures compared with traditional teaching. Barker and Millar (2000) found in a longitudinal study on the Salters Advanced Chemistry course that a gradual introduction and revisiting of chemical ideas in different contexts appeared to have a significantly positive impact on the learning outcomes of a high proportion of students. Although these findings indicate that context-based courses can facilitate the development of students' conceptual coherence, the underlying mechanisms that describe how this development proceeds still need to be unraveled (Gilbert, Bulte, & Pilot, 2011; Pilot & Bulte, 2006). Therefore, identifying the principles underlying learning-teaching (LT) activities that foster the development of students' conceptual coherence is regarded as one of the major challenges in research on context-based science education.

In response to this challenge, we have adopted a design research approach (McKenney & Reeves, 2012; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). In this paper, we describe a case study that focuses on the design and evaluation of a context-based lesson sequence. We pay specific attention to a two-step evaluation procedure, which provides profound insight into the teaching and learning processes. First, we evaluate the lesson sequence based on its practicability and determine the extent to which the lesson sequence is conducted as intended. The methodological approach is built on a previous study within our research project in which we described how a research scenario was used to compare the intentions of the design with the way in which it was enacted in the classroom (Ummels, Kamp, de Kroon, & Boersma, 2015). Second, we evaluate the lesson sequence based on its effectiveness. This evaluation is built on another study within our research project in which we showed that determining changes in how students mentioned concepts and their interconnections (propositions) from a reference concept map during a lesson sequence gives information about students' development of conceptual coherence (Ummels et al., 2013). We thus determined to what extent students' conceptual coherence has developed.

The learning goals of the lesson sequence are aimed at a conceptual framework in the domain of biology that is complex and difficult to learn and teach: transformations of forms

of energy and carbon-substances in photosynthesis, cellular respiration, and biosynthesis (Amir & Tamir, 1990; Brown & Schwartz, 2009; Cañal, 1999; Lin & Hu, 2003; Mohan, Chen, & Anderson, 2009). This conceptual framework is embedded in contexts that are related to the social and scientific debate on the impact of protein-rich food production on the environment (McMichael, Powles, Butler, & Uauy, 2007). A context is defined as a representation of "an authentic community of practice within society" wherein students, supported by the teacher, work collaboratively on relevant tasks in a problem-centered way for a sustained period (Gilbert et al., 2011). During these tasks they are expected to deal with biological concepts and to establish relationships between these concepts. Our aim is to explore how students' conceptual coherence develops to contribute to optimizing context-based science lesson sequences in general. Such information will be valuable for educational researchers who examine how context-based lesson sequences work in practice, educational designers of similar context-based lesson sequences in similar settings, and teachers who conduct these lesson sequences. Consequently, we address the following two research questions:

1. How does conceptual coherence develop for students during a context-based lesson sequence?
2. How do context-embedded LT activities influence the development of conceptual coherence?

THEORETICAL FRAMEWORK

Conceptual Learning

This research is built on the theory that concepts are fundamental units of knowledge and that learners do not store these concepts as isolated bits of information but form connections between concepts (Ausubel, 1968). Conceptual learning occurs when a new concept is assimilated actively and meaningfully in someone's cognitive structure. This is also one of the basic assumptions of constructivist approaches on learning (Ogborn, 1997). This theory implies that a new concept is connected to one or more relevant "existing" concepts. This process proceeds when more elaborate connections are established between two or more concepts (Mintzes, Wandersee, & Novak, 2005). So, the more connections (or cognitive pathways) established between concepts, the greater the chance of retrieving these concepts. Studies on expert learning showed that an easy retrieval of concepts is supported by a systematic or hierarchical organization of concepts in cognition because memory easily "travels down" these well-worn pathways (Fisher, 2001). Moreover, a more coherent organization of conceptual knowledge enables experts to represent and solve new problems more successfully than novice learners (Bransford et al., 2000a). One characteristic of this coherent conceptual organization is the presence of inclusive or core concepts that structure other, often more descriptive, concepts.

One way to represent the (intended) conceptual relationships within a learner's cognitive structure is to make use of concept maps (Novak & Cañas, 2008). A concept map is a graph consisting of concepts connected by labeled lines. Two of these connected concepts within a concept map are known as a proposition. Therefore a proposition can be regarded as the smallest unit of coherent conceptual knowledge (Mintzes et al., 2005).

Context-Based Biology Education

There are many variations in the definition of context-based approaches in science education (King & Ritchie, 2012). Here, we focus on a specific form of context-based

education in the domain of biology that is currently a focus of Dutch educational reform: the concept-context approach (Boersma et al., 2007). This approach is rooted in cultural historical activity theory (Vygotsky, 1987). In this approach, contexts are defined as representations of existing scientific, professional, or real-life practices. To engage students in contexts, suitable social practices need to be transformed for classroom use in such a way that students experience them as relevant. This transformation happens when students recognize the perspective of participants of such social practices or imagine themselves as these participants. From this perspective, they perform goal-oriented activities in which they enter into a *cognitive apprenticeship* with the teacher. The teacher demonstrates how a context and its activities might be interpreted, being aware of individual students' *zones of proximal development*. Finally, relevant concepts are summarized and their meaning in the given context considered. The different, often subtle, meanings of a concept in two or more contexts can be compared. This comparison facilitates the process of recontextualization, which means that a concept is transferred from one context to another (Van Oers, 1998).

Learning Teaching Activities

This research focuses on conceptual learning that takes place during LT activities that are embedded in contexts. We define LT activities as delimited educational units that consist of an introduction phase, an action phase, and reflection phase in which students and teacher perform activities. Literature provides evidence that several LT activities (or elements that can be integrated in LT activities) are associated with conceptual learning. One of these LT activities is writing. It was shown that writing activities about real-life topics improve the abilities of students to integrate concepts and apply concepts to real-world problem solving (Keselman, Kaufman, Kramer, & Patel, 2007). During the writing process, students are stimulated to brainstorm, which activates associations among concepts that are stored in long-term memory (Galbraith, 1999). Moreover, writing prompts students to organize their conceptual knowledge and to express relationships between concepts when formulating sentences.

There are strong indications that constructing concept maps, another LT activity, is associated with increased knowledge retention and transfer (Nesbit & Adesope, 2006). This can be explained because the brain organizes concepts in a parallel way (Novak & Cañas, 2008). Concept mapping appears to be particularly useful in assisting students to understand the interconnectedness of complex biological relationships (Kinchin, 2011). Other visualization tools, such as flow charts, also seem to help improve student comprehension and learning (Davidowitz & Rollnick, 2001; McCabe, 2011). They can reveal interrelationships and connections within knowledge and can therefore be seen as a tool to make students' conceptual thinking visible (Ritchhart, Turner, & Hadar, 2009). Finally, teacher–student conversations play an important role in developing students' conceptual understanding. The kinds of questions the teacher asks and the way in which the teacher articulates these questions can stimulate students to construct new relations between concepts (Chin, 2007).

Design Principles

We used four design principles, specific to the concept-context approach, to steer the design of a lesson sequence and in particular the structuring of LT-activities within contexts. Following Van den Akker et al. (2006), we defined these design principles as theoretically and empirically grounded constructs, linking strategies with intended pedagogic effects. In the following formulations, where possible, we refer to findings of a previous case study within our research project (Ummels et al., 2015).

Building upon concepts students are expected to be familiar with. In line with constructivist approaches to learning and teaching, attention to previously acquired (conceptual) knowledge is a prerequisite for learning new concepts (Novak, Mintzes, & Wandersee, 2005). This implies that when a context is introduced in the classroom the initial focus should be on concepts with which students are expected to be familiar from personal life or prior education. These concepts can function as "stepping stones" to introduce new concepts. The previous case study showed that the questions asked by the teacher are important in scaffolding students to widen their thinking from the concepts they are familiar with to new concepts.

Focusing on core concepts. Studies on expert learning have shown that experts organize their knowledge around core concepts that guide their thinking (Novak & Cañas, 2008). Introducing core concepts in a drip-feed manner in different contexts and constantly reinforcing them in different ways seem to be fruitful learning and teaching strategies (Barker & Millar, 2000). Our previous case study also provided indications that a problem-posing approach (Klaassen, 1995), in which a context-related problem is solved in a guided step-by-step fashion, could be a useful strategy for creating a motive to focus on core concepts within a context.

Stimulating students to interconnect concepts. When students are stimulated to interconnect concepts actively and frequently it is expected that their cognitive connections between these concepts will be reinforced (Fisher, 2001). In the previous case study, it became clear that LT activities embedded in contexts, in which students had to link concepts meaningfully, challenged them to make their conceptual thinking visible and to discuss the correctness of propositions with each other and with the teacher.

Reflecting on conceptual relationships within a context. Learning to recontextualize concepts from one context to another is assumed to enhance conceptual coherence (Van Oers, 1998). Recontextualization requires that students are supported to reflect on the interrelationships of concepts within a context (Wierdsma, 2012). The previous case study showed that if there was no need for students to reflect on propositions within a context this resulted in a teacher-guided noninteractive recapitulation of concepts and propositions.

Although there is no prescribed order in which these design principles should be applied it seems natural that the first design principle is elaborated at least at the beginning of a context and the fourth design principle at least at the end of a context.

METHODS

Reference Concept Map

The concepts and propositions students were expected to learn during the lesson sequence were presented in a concept map. Because each proposition that students mentioned could be pointed out in this concept map, we called it the *reference concept map*. This reference concept map was used to guide the design and as a tool to assess improvements in mentioning propositions during the course of the lesson sequence.

The concepts to be learned were selected from two biology textbooks for upper secondary education and the national Dutch exam standards (CvE, 2009) and were based on the question of which photosynthesis-related concepts are important to teach to 10th-grade biology students in senior general secondary education. Next, we conducted an analysis of the relevant literature about learning and teaching photosynthesis and other metabolic processes. We identified the following three main problems:

- Students do not understand cellular "processes." They consider photosynthesis and cellular respiration as exactly opposite processes or purely as "gas exchanging" processes (Cañal, 1999; Kose, Usak, & Bahar, 2009).
- Students are not used to seek explanations at the cellular or subcellular level of biological organization when they are asked to explain observable natural phenomena (Flores, Tovar, & Gallegos, 2003; Songer & Mintzes, 1994).
- Students are not able to link the living world to the nonliving world. They often do not grasp the idea that in living things energy can be captured, transferred, or released and that chemical elements (like carbon) can be transformed in a cyclic way from one molecule to another (Amir & Tamir, 1990; Lin & Hu, 2003; Mohan et al., 2009).

This last problem is not surprising considering that textbooks do not convey the idea that the metabolic processes (photosynthesis, cellular respiration, and biosynthesis) in living things are instances of matter and energy conservation and transformation (Roseman, Linn, & Koppal, 2008). After discussions with researchers in the field of ecology and upper secondary biology teachers, the following two guidelines for the construction of the reference concept map were devised:

- The three metabolic processes of photosynthesis, cellular respiration, and biosynthesis need to be related to each other at the cellular level.
- Each process should present how matter (with a focus on carbon-containing substances) and forms of energy (light energy, chemical energy, heat energy, and energy for cellular work) are converted.

On the basis of those two guidelines, we defined the relationships between concepts, which resulted in four groups of propositions focused on the *core* concepts of photosynthesis, cellular respiration, biosynthesis, and energy. The propositions related to energy refer to the release of heat energy or energy for cellular activity from chemical forms of energy. Figure 1 shows the reference concept map containing all these propositions.

Overview of Lesson Sequence

To guide the selection of authentic social practices which can be transformed into contexts, we chose a socioscientific topic: the environmental impact of producing meat and other protein-rich food products (Tytler, 2005). After making an inventory of social practices that related to this topic, we focused on those that (1) had the potential to provide a framework for the setting of "focal events: important or typical events that draw the attention of learners while remaining imbedded in its cultural setting" (Gilbert et al., 2011); (2) could be transformed into contexts that covered as many concepts and propositions from the reference concept map as possible; and (3) could be interlinked by a storyline. Eventually, we chose three practices that could be interconnected by a guiding question.

The first context is representative of a family-life practice: a family discussing whether to become vegetarian or not from a biological perspective. In this discussion, vegetarianism is linked to the agricultural journey of meat and meat substitutes. Students role-play a discussion about meat consumption. It is assumed that students recognize that such discussions could reflect their own situations. This context ends with the question: *Will we still be allowed to consume meat in the future (from a biological perspective)?* In this context, the following concepts from the reference concept map are introduced: carbon dioxide, proteins, and energy for cellular activity (Figure 1). So far, no propositions from the reference concept map are introduced.

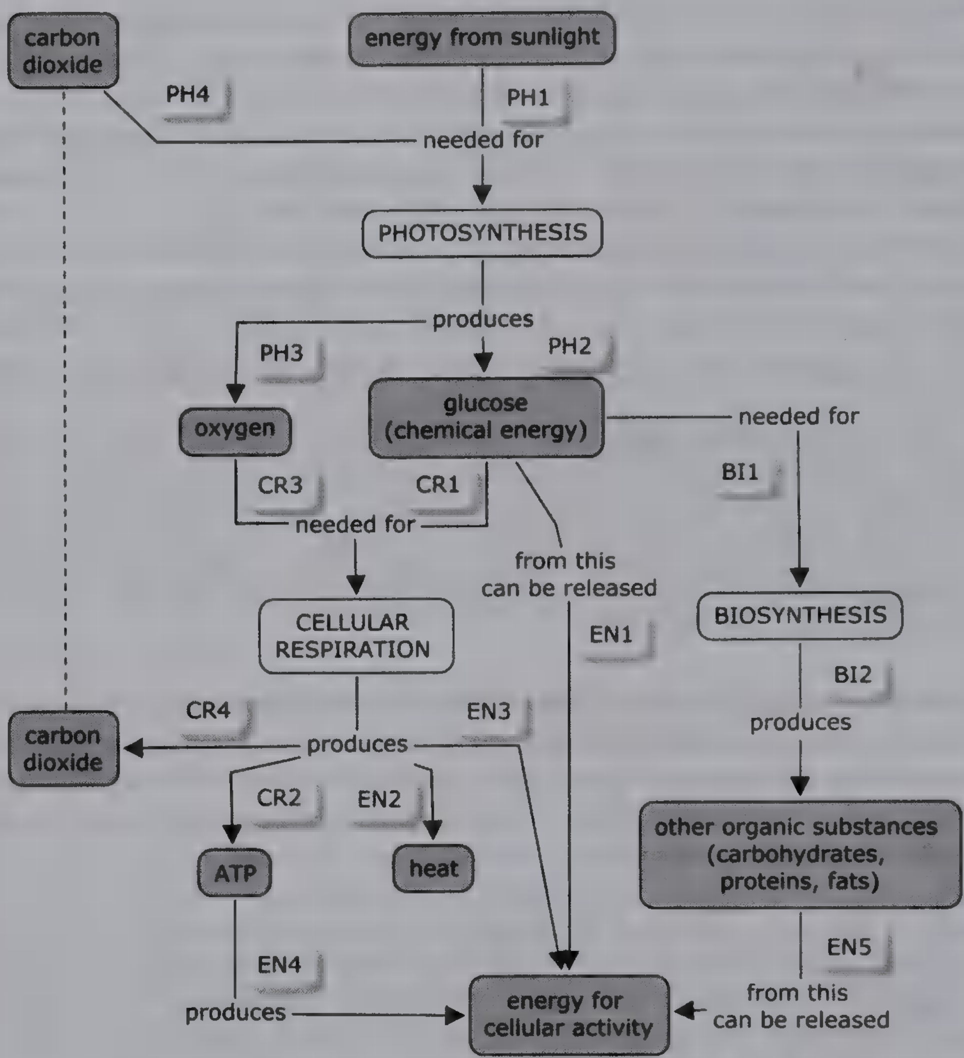


Figure 1. Reference concept map. The relationships between the three metabolic processes (white boxes) with an emphasis on transformations of forms of energy and matter (gray boxes) are indicated with proposition codes. Four groups of propositions are distinguished that are related to the core concepts of photosynthesis (codes: PH1–4), cellular respiration (codes: CR1–4), biosynthesis (codes: BI1–2), and energy (codes: EN1–5).

The second context is representative of the professional practice of an environmental advisor. The task of this advisor is to describe the impact of the agricultural production of meat and meat substitutes. Because students may not be familiar with this profession, a chart is developed to visualize the context. This graphic visualization is used to introduce the relation between the greenhouse effect and food production, with a focus on carbon dioxide emissions during the production of various plant- and animal-based protein-rich food products. Figure 2 shows the graphic visualization of the context in which the role of the environmental advisor (in front) in the protein-rich food production chain has been displayed. Protein-rich food is specified as meat for beef burgers and soya for vegetable burgers. Other participants involved are a consumer, a cattle farmer, a crop (soya) farmer, and an agricultural researcher.

Students use the graphic visualization of the context to indicate where in the production chain there is a release and an intake of carbon dioxide. Figure 3 shows the arrows students are expected to draw in the graphic visualization of the context. Furthermore, the graphic visualization of the context is used to link carbon dioxide to two processes in cells: cellular respiration and photosynthesis. Specific attention is given to the propositions PH4 and CR4

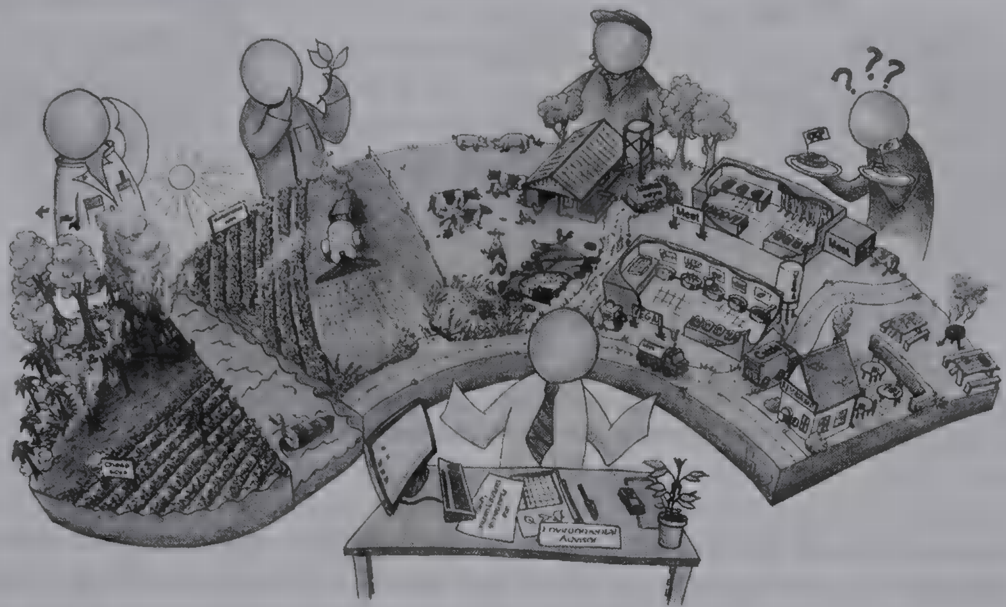


Figure 2. Graphic visualization of context. In front the environmental advisor who has to deal with the impact on the environment during the production chain of protein-rich food products. Other participants involved (anticlockwise): a consumer, a cattle farmer, a crop farmer, and an agricultural researcher.

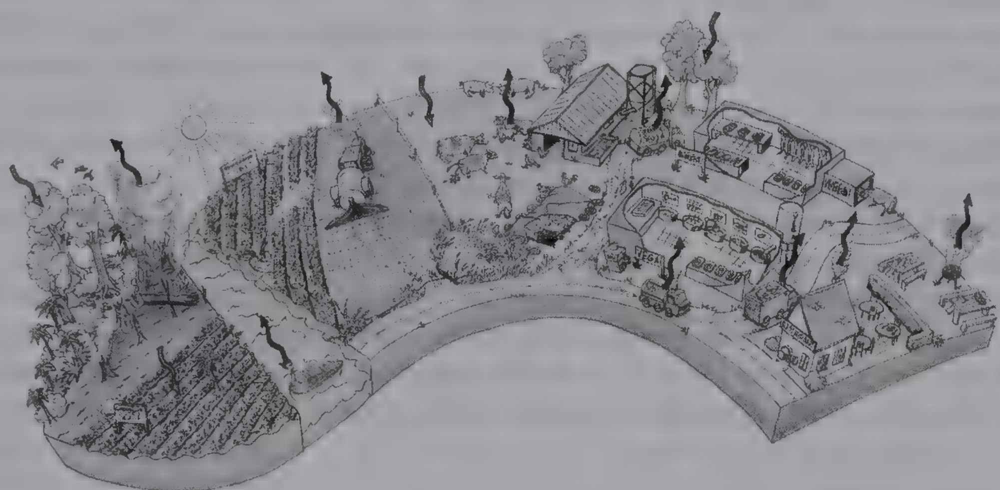


Figure 3. How the graphic visualization of the context is used to indicate where in the food-production chain there is an intake and a release of carbon dioxide.

(Figure 1). To explain this, the teacher uses the graphic visualization of chloroplasts in plant cells and mitochondria in animal and plant cells (Figure 4). In this explanation, he emphasizes the propositions PH4, PH1, PH2, CR1, CR4, EN2, and EN3 (Figure 1). The use of these graphic visualizations is the first LT activity we focus on in this article. At the end of the second context, students have to write advice destined for a public information association named "Consumer and Environment" from the perspective of the environmental advisor. This is the second LT activity we focus on in this article. The guiding question is specified as *will we still be allowed to consume meat in the future with regard to carbon dioxide emissions?* It is expected that students mention in this draft version at least the propositions that include carbon dioxide: CR4 and PH4 (Figure 1).

The third context is representative of the scientific practice of researchers in agriculture. These researchers study the production of soya plants in relation to the production of animal

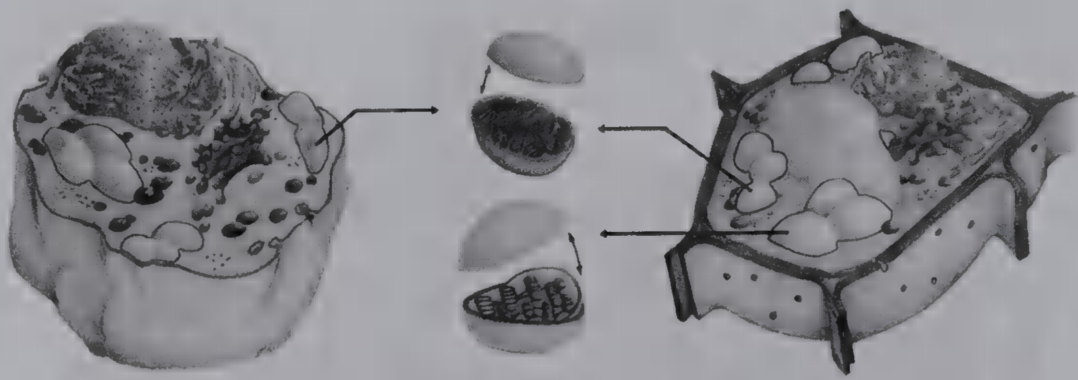


Figure 4. Graphic visualizations of an animal cell (left) and a plant cell (right), mitochondrion (top, center) and chloroplast (bottom center) used by the teacher to focus on cellular respiration and photosynthesis.

and plant proteins in the food we eat. Students weigh in pairs all soya beans from one soya plant and calculate how much larger surface of agricultural land is needed for the production of 1 kilogram of proteins in meat compared to 1 kilogram of proteins in soya plants. To explain this difference, the teacher provides students with knowledge of the metabolic process of biosynthesis (and relates this to photosynthesis and cellular respiration). The propositions BI1 and BI2 (Figure 1) are introduced. Furthermore, students construct concept maps about how the three metabolic processes are involved in protein production in both plant and animal cells. They are expected to establish the propositions PH1 to PH4, CR1 to CR4, BI1 and BI2, and EN1, EN2, and EN4 (Figure 1). This is the third LT activity we focus on in this article.

Finally, the lesson sequence returns to the perspective of the environmental advisor. Students perform a second writing task in which they use their earlier written text and information from the manual to write a final version of the advice to the public information association "Consumer and Environment." For a correct and complete explanation, it is expected that students use the propositions PH4, PH2, CR1, CR2, CR4, EN2, and EN3 (Figure 1). The guiding question is now extended to *will we still be allowed to consume meat in the future with regard to both carbon dioxide emissions and the use of agricultural land?* This is the second part of the second LT activity we focus on.

Performance of the Lesson Sequence

The lesson sequence was conducted in a 10th-grade biology class of 29 students, aged 15–16, in senior general secondary education. This type of education prepares students for studies at a University of Applied Science. To assess the abilities of these students, we compared their average biology grades for 1 year with those of students from a parallel class. The former class, on a scale of 1–10, scored an average grade of 6.17 ($SD = 0.66$), whereas the latter class scored an average grade of 6.35 ($SD = 0.64$). This is not significantly different at a 95% confidence interval: $t(58) = -1.12$; $p = .269$. The school was located in a semirural area in the east of The Netherlands. The teacher, who specializes in teaching students in senior general secondary education, had about 15 years of teaching experience and is a highly competent biology teacher. In advance of each lesson, the first author and the teacher discussed the intended LT approach. To support the teacher during the lessons, there were digital presentations that provided him with information: for example, the questions he could ask in a classroom conversation. Instructing and preparing the teacher as accurately as possible minimizes the influence of undesirable actions on students' learning processes. Because the school participated in a pilot project for the implementation of

biology education based on the concept-context approach, students were familiar with this type of education. The lesson sequence was conducted in 10 consecutive lessons within a period of 3 weeks. To limit the variation in students' time spent on the lesson sequence, students were not allowed to take their manual home during this 3-weeks period. Only for the final test, they were allowed to study the manual at home.

Research Scenario and Evaluation

A research scenario predicts and theoretically justifies in detail the expected LT process and why it is expected to happen in that particular way (Lijnse & Klaassen, 2004). In this study, we constructed a research scenario that included a stepwise description of context-embedded LT activities that were expected to contribute to students' conceptual coherence. These expectations were based on a pedagogical analysis of the content to be taught (see the section Reference Concept Map) and a previous case study on concept-context-based education (Ummels et al., 2015). In this case study, we examined why LT activities were effective from the students' perspective. Table 1 demonstrates this research scenario. For the purpose of this article, only three LT activities are presented here. They are called *using graphic visualizations, writing,* and *concept mapping*. Data from video recordings in the classroom were collected and transcribed verbatim to evaluate these LT activities on practicability. This evaluation focused on the degree of correspondence of the intended steps and what actually happened in classroom practice.

One of the following scores was assigned to each step: positive (+) when the step was fully recognized in the transcripts; negative (–) when the step was not recognized in the transcripts; or intermediate (±) when the step was partially recognized in the transcripts. A second rater (second author) followed the same procedure. The level of agreement between the raters appeared to be high (Cohen's kappa = 0.87). Observational remarks on intermediate and negative scores were noted. The scores were used to interpret how characteristics of the design might have contributed to students' learning processes. The assumption is that when essential steps are missing or not performed as intended this might interrupt the development of conceptual coherence.

The first four steps (1.1–1.4) of the first LT activity (*using graphic visualizations*) were scored positively. Therefore, it was expected that students could relate carbon dioxide to cellular respiration (mentioning proposition CR4; see Figure 1) and photosynthesis (mentioning proposition PH4). The teacher made some incorrect formulations, however, about the concept of energy during the reflection phase when chemical equations were discussed (Step 1.5). For example, the teacher said: "Here is glucose and during cellular respiration this matter is partly transformed into energy." This might have resulted in a problematic understanding of transformations of matter and forms of energy.

Concerning the second LT activity (*writing*), the teacher introduced and instructed the writing assignment as intended (Steps 2.1–2.3). Students asked the teacher many questions during the brainstorm phase of the writing assignment (Step 2.4) and when they were supposed to write individually (Step 2.5). Obviously, they experienced problems in applying conceptual knowledge when giving explanations. Moreover, interaction that prompted students to reflect on conceptual relationships did not really take place (Steps 2.6 and 2.7). To help students to write the final advice, the teacher started with an introduction (Step 2.8) that missed the essential point, which had to be explained in the advice: the impact of consuming meat on both carbon dioxide emissions and the use of agricultural land. Students wrote their texts individually and used the manual as intended (Step 2.9). Furthermore, final reflection (Step 2.10) did not take place.

TABLE 1
Research Scenario for Three Context-Embedded LT Activities of the Lesson Sequence

	Specification of LT Activities	Score	Observational Remarks on Intermediate and Negative Scores
LT Activity 1: Using graphic visualizations (part of Context 2)			
1.1	Teacher uses graphic visualization of context (Figure 2) to make clear that environmental advisor studies carbon dioxide emission in food production chain.	+/+	
1.2	Teacher gives instruction to draw arrows in graphic visualization of context indicating carbon dioxide exchange.	+/+	
1.3	Students discuss in pairs and draw arrows in graphic visualization of context (Figure 3).	+/+	
1.4	Students check correctness of arrows in graphic visualization of context.	+/+	
1.5	In reflection phase, the teacher uses graphic visualizations of contexts (Figures 2 and 3) in combination with graphic visualizations of plant and animal cells (Figure 4) to explain the link between cellular respiration and photosynthesis mentioning the propositions PH1, PH2, PH4, CR1, CR4, EN2, and EN3.	+/±	The teacher did not formulate the concept of energy correctly.
LT Activity 2: Writing (part of Contexts 2 and 3)			
	Part 1: Draft version		
2.1	Teacher legitimizes why an environmental advisor writes advice.	+/+	
2.2	Teacher indicates that the texts should make clear how carbon dioxide emission in the production of protein-rich products of a vegetable origin is lower than for those of animal origin.	+/+	
2.3	Teacher instructs on steps in writing process.	+/+	
2.4	Students brainstorm in pairs: They discuss the line of reasoning and go through their manual.	±/±	Many procedural questions from students indicated that they did not seem to understand the assignment.
2.5	Students individually write a draft version.	±/±	Teacher's support still needed.
2.6	In reflection phase, the teacher asks students to point out essential points in argumentation. There is an emphasis on the use of propositions CR4 and PH4.	±/±	Reflection was limited to just a presentation of a slide.

(*Continued*)

TABLE 1
Continued

Specification of LT Activities	Score	Observational Remarks on Intermediate and Negative Scores
Part 2: Final version		
2.7 Students read each other's written products and give feedback on established propositions.	–/–	Reacting to each other's products did not occur. Time was lacking.
2.8 Teacher indicates that the text of the final version of the advice should make clear how it is possible that both carbon dioxide emissions and use of agricultural land for the production of protein-rich products of a vegetable origin can be lower than those of animal origin.	±/±	Teacher recapitulates activities in prior lessons but does not mention the main focus of the final version.
2.9 Students individually write a final version and use manual.	+/+	
2.10 In reflection phase, the teacher points out essential points in argumentation. In this argumentation there is an emphasis on the propositions PH4, PH2, CR1, CR4, EN2, EN3, BI1 and BI2.	–/–	This step was not recognized. There was no time
LT Activity 3: Concept mapping (part of Context 3)		
3.1 Teacher explains goal: shows in two concept maps how plant cells and animal cells produce proteins.	+/+	
3.2 Teacher instructs on steps in mapping process and gives each student in group a responsible role.	+/+	
3.3 Students construct maps and interact about connections between concepts to be made.	+/+	
3.4 Teacher gives each group feedback and asks students to explain connections they have made.	+/+	
3.5 Students compare own concept maps with those of other groups.	+/+	
3.6 In reflection phase, the teacher challenges students to use concept maps to explain why proteins are produced more efficiently in plant cells than in animal cells, referring to transformations of forms of energy. The following propositions are mentioned: PH1-PH4, CR1-CR4, EN1, EN2, EN4, BI1 and BI2.	±/±	There was limited time, the teacher recapitulates with limited interaction.

Each step of an LT activity was scored with regard to how well it was performed compared with the intended performance. The proposition codes refer to Figure 1. The scores of two researchers (divided by /) are presented as follows: a positive score (+) when a step was observed as intended; a negative score (–) when a step was not observed as intended, and an intermediate score (±) when a step was partially observed as intended. Observational remarks on intermediate and negative scores are presented in the last column.

The first five steps of the third LT activity (*concept mapping*) were performed as intended. Students looked focused, worked cooperatively in groups when trying to establish propositions, and followed the procedural steps to construct a concept map (Steps 3.1–3.3) and to check with other groups (Step 3.5). Moreover, the teacher interacted with each group and gave feedback (3.4). Only the last step (3.6), in which it was intended that the teacher would ask students to use the concept map when giving explanations, was not recognized. Here the teacher taught by telling and did not interact with students. Possibly, the teacher did not recognize the propositions that were not fully understood by students.

Data Collection and Analysis

Various data sources were collected by multiple means and at different points to shed light on how this lesson sequence improved the development of students' conceptual coherence. These data sources consisted of video recordings of all lessons, written responses on a pretest, a posttest, and a final test, semistructured interviews with four students and the teacher, and short written evaluations of all students after each lesson (postlesson evaluations). Moreover, we collected two "naturalistic" data sources (products of the lesson sequence itself): concept maps as products of group work and the draft and final texts of the writing assignment. Each data source provides a different insight into students' learning processes. Therefore, the triangulation of these data was used to describe how students' conceptual understanding developed during the lesson sequence.

Pretest, Posttest, and Final Test. Identical pre- and posttests were performed, consisting of explanatory tasks and defining tasks. In the explanatory tasks, students had to predict what would happen with a number of shrimp and green algae that were trapped for 1 year in a sealed ecosphere, containing water and with unrestricted sunlight. They also had to predict what would happen with a mouse and a green plant that were individually trapped in a sealed container with unrestricted water and sunlight. This task was derived from an empirical study on the understanding of cellular respiration and photosynthesis of college-level biology students (Songer & Mintzes, 1994). Our intention with these tasks was to confront students with inconsistencies in their thinking and persuade them to formulate plausible solutions to biological problems. Correct responses should refer to photosynthesis as *a process to capture energy* and to cellular respiration as *a process to release energy*. Moreover, students had to explain the relationship between animals and plants in terms of oxygen and carbon dioxide release and intake. This task was inspired by a study of Tamir and Amir (1990) who showed that 11th- to 12th-grade students often think that cellular respiration only takes place in animals and that photosynthesis is the opposite process of cellular respiration. Most students do not know that the processes are complementary.

In the defining tasks, students were asked to formulate a definition of the following concepts from the reference concept map: cellular respiration, biosynthesis, photosynthesis, organic substances, and adenosine triphosphate (ATP). Moreover, they had to relate some given concepts from the reference concept map: photosynthesis, sunlight, carbon dioxide, heat, cellular respiration, ATP, glucose, water, and oxygen. We assumed that the propositions mentioned in these defining tasks could be obtained from merely reproductive learning. The propositions mentioned in the explanatory tasks, in which no core concepts were given, are more likely to be the result of meaningful learning (Mintzes et al., 2005). The posttest was performed immediately after Lesson 10. Students responses on the pretest prediction tasks were returned to them, and they were asked to check these responses and to indicate

adjustments. For the defining tasks, students were also asked to adjust their definitions given in the pretest and to add definitions they could not formulate in the pretest.

In the final test, the responses to 10 questions were analyzed: eight context-oriented open questions in which students had to give explanations, comparable with the explanatory tasks in the pre- and posttest, and two multiple-choice questions. These were selected because a correct choice required students to relate four propositions to one another. In one question students were asked to relate the core concepts of photosynthesis and cellular respiration, and in the other question they were asked to relate photosynthesis and biosynthesis. If students are able to relate a combination of two metabolic processes, this indicates a high degree of conceptual understanding.

For all three tests, we coded the propositions from the reference concept map that were expected in correct responses. For each proposition, we quantified how many students were able to mention it at least once in each test exactly as described in the reference concept map. In line with previous work, we assumed that mentioning propositions is an indicator of the degree to which conceptual coherence has developed (Ummels et al., 2013). A second rater followed the same procedure for five randomly selected students. In this procedure, students' concept maps and written products collected during the execution of the lesson sequence were also included (see the section Products of Concept Mapping and Writing Assignments). The level of agreement between the raters was high (Cohen's kappa = 0.97). Apparently, the codebook that described which remarks had to be scored as a correct proposition was unambiguous. The results obtained from responses on the defining and explanatory tasks were compared between pretest, posttest, and with the open context-oriented questions in the final test. Statistical analysis was conducted between the identical pre- and posttests. Because of the relatively low numbers of students, we conducted a nonparametric sign test to determine whether there was a significant increase between the two tests in the numbers of students who mentioned propositions correctly. The results of the two multiple-choice questions were compared to provide additional information on students' abilities to establish combinations of propositions. These results should be treated with considerable caution because the validity of multiple-choice questions is often doubted (Mintzes et al., 2005).

Semistructured Interviews. Semistructured interviews (Southerland, Smith, & Cummins, 2005) were conducted with four students individually after every two lessons and after the pre- and the posttest. Each student was interviewed six times. An interview lasted about 30 minutes and had a similar structure. These students (two males, aged 16 and 17, and two females, aged 15 and 16) were selected by the teacher on the basis that they represented different learning styles and were cooperative. We compared these four students' grades for biology during 1 year with the grades of the other students in the class. The mean grades of the four students were on average –0.67 standard deviations lower than the mean grades of the whole class ($n = 30$ students). Standardized differences of these four students ranged from –1.91 (student 26) to +0.51 (student 11). This shows that these four students were not outliers. The research scenario was used to formulate appropriate probes and follow-up questions to gain a complete understanding of the interviewees' views. Video recordings of these interviews were transcribed verbatim and analyzed by close reading and highlighting passages that indicated how students' learning processes had occurred with respect to understanding and interconnecting core concepts from the reference concept map. We focused on remarks that included single propositions (e.g., photosynthesis produces glucose) and combinations of propositions connecting two or more core concepts (e.g., photosynthesis produces glucose which is needed for cellular respiration and biosynthesis).

This was followed by axial coding, which allowed the information to be clustered and summarized. Students' ideas about the usefulness of each of the three LT activities were also inventoried. Three semistructured interviews were conducted with the teacher. The transcripts of these interviews were analyzed by looking for passages that indicated how the teacher perceived that the LT activities influenced students' development of conceptual coherence, either positively or negatively.

Products of Concept Mapping and Writing Assignments. Two data sources were collected from the "naturalistic setting": the texts of the writing assignment and concept maps as products of group work. The two texts of the writing assignment were analyzed by coding the propositions from the reference concept map and counting the numbers of students who mentioned each of these propositions. In the draft version, students had to link consumption of meat and other protein-rich food products to carbon dioxide emissions. In the final version, students had to explain how consuming meat is also related to use of agricultural land. The concept maps were analyzed by coding each proposition from the reference concept map, indicated by an arrow and a label connecting two concepts. For each proposition, we counted how many times it was recognized in nine concept maps. Identifying which and to what extent students were or were not able to establish propositions provides information about students' development of conceptual coherence.

RESULTS

Changes in Mentioning Propositions Before and After Lesson Sequence

Figure 5 shows the numbers of students who mentioned each proposition in defining and explaining tasks in the pretest, posttest, and final test. We discuss the results with respect to the propositions related to each of the four core concepts: photosynthesis, cellular respiration, biosynthesis, and energy.

In the pretest, more students were able to mention propositions in relation to the core concept of photosynthesis than the other core concepts. This is not surprising because the chemical equation of photosynthesis was taught in an earlier module. In the defining tasks, many students wrote down this chemical equation when asked to define photosynthesis. Fewer students mentioned propositions related to photosynthesis in the explanatory tasks. This might indicate that establishing propositions in new situations is more difficult than reproducing propositions. There were only two students (no. 5 and no. 16) who mentioned proposition PH2: "algae produce glucose (as a food source for shrimps) by photosynthesis." Although there were more students who stated: "algae produce food for shrimps" without mentioning glucose, this was not scored as a correct proposition. Ten students mentioned the propositions PH3 and PH4 in one sentence: "Algae use sunlight in photosynthesis to turn carbon dioxide into oxygen." This indicates that these students consider photosynthesis mainly as a gas exchanging process. This problem is also reported in the literature (Amir & Tamir, 1990; Cañal, 1999).

Interviews with four students (no. 11, 13, 21, and 26) conducted after the pretest made clear that their prior knowledge contained the idea that photosynthesis is a "mysterious" process taking place in plants and that plants need sunlight, use carbon dioxide, and produce oxygen. One of the students (no. 21) said, "Plants perform all kinds of tricks like photosynthesis." These students often reasoned from the perspective of the organism, focusing on what a plant needs and produces. The process of cellular respiration seemed to be largely unknown to these students. Each of the four students mentioned that animals

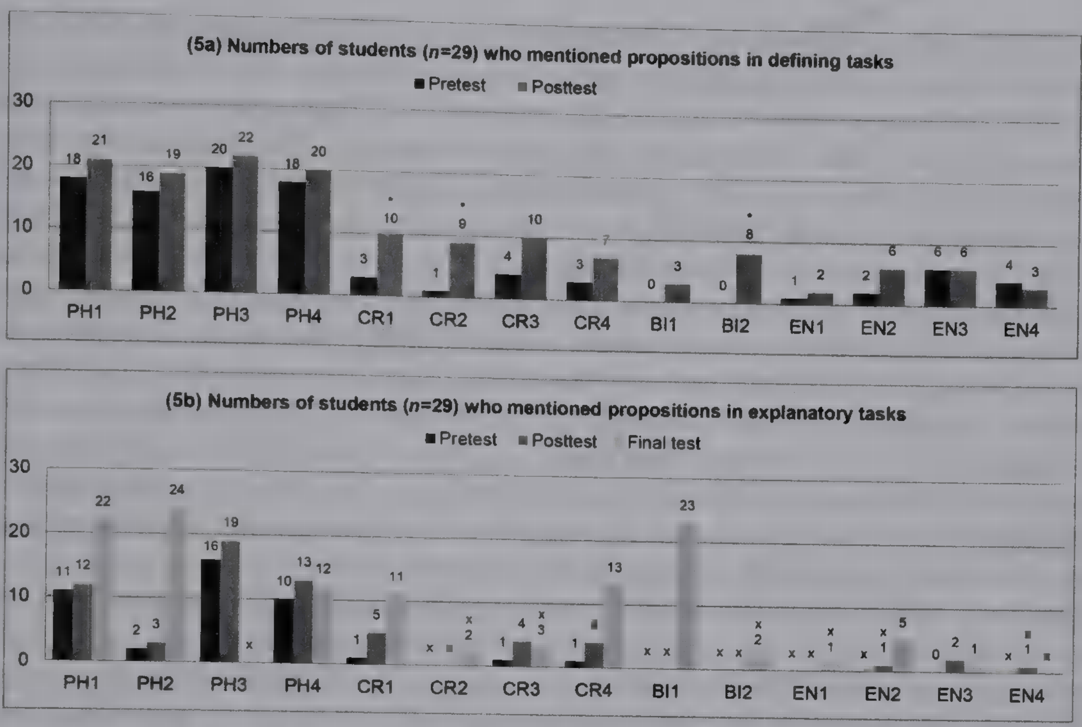


Figure 5. Numbers of students who mentioned propositions related to the concepts of photosynthesis (PH1–4), cellular respiration (CR1–4), biosynthesis (BI1–2), and energy (EN1–4) in defining tasks at pre- and posttest (a) and at the explanatory tasks at pretest, posttest, and final test (b). The asterisks (*) in Figure 5a indicate significant differences between pre- and posttest scores. The final test was not included in statistical analysis. The crosses (x) in Figure 5b indicate propositions that were not evoked and therefore not expected. EN5 was in none of the tasks evoked and is excluded from these figures. The codes refer to the propositions shown in the reference concept map as presented in Figure 1.

need oxygen but could not explain this by referring to cellular respiration. Student no. 13 said, "I think oxygen is needed for blood circulation." Only student no. 26 said, "I think that plants also use a bit of oxygen but I'm not sure." All four students had problems shifting their locus of explanation from the organism level of organization to the cellular or subcellular level. From the defining tasks, it became clear that none of the students could describe biosynthesis. Furthermore, the students were not able to describe how forms of energy could be converted into one another—for instance, that light energy could be captured as a form of chemical energy (glucose) and that this chemical energy could be transformed into energy for cellular work. Student no. 13 said, "Energy from sunlight is turned into oxygen." None of the four students were able to describe any transformations of forms of matter or energy during one of the metabolic processes. This phenomenon was also recognized by Lin and Hu (2003).

Total test scores were computed for the number of correctly mentioned propositions in both the pre- and the posttest. The sign test showed a significant increase in the total number of propositions that were mentioned correctly in both defining tasks (16 positive and three negative differences and 10 ties, $p = .004$) and explanatory tasks (10 positive differences and no negative differences and 19 ties, $p = .002$) between pre- and post-test. For each proposition separately, the only significant increase between pre- and posttest was found for the propositions CR1 (seven positive differences and 22 ties, $p = .016$), CR2 (nine positive and one negative differences and 19 ties, $p = .021$) and BI2 (eight positive differences and 21 ties, $p = .008$) in the defining tasks. No significant differences between pre- and posttest were found for the other propositions (for all: $p > .05$).

Although students adjusted their responses, they often only mentioned concepts they had not mentioned before (mainly concepts related to cellular respiration) and they did

not establish any new propositions. There were only three or four students who mentioned three propositions concerning cellular respiration—CR1 (four), CR3 (three), and CR4 (four)—in the explanatory tasks, but they did not mention these propositions in the pretest. For example, one of the students (no. 10) said in the pretest, "The plant (in the sealed container) is going to die because it needs carbon dioxide to live. This carbon dioxide is transformed into oxygen and when it runs out of carbon dioxide it dies." The student corrected this response in the posttest into the following: "The plant needs carbon dioxide for photosynthesis, which also requires water and sunlight. This process produces oxygen and glucose which the plant needs for cellular respiration. From this process carbon dioxide, water and sunlight are released." Although this student related the process of photosynthesis to cellular respiration, he seemed unfamiliar with the concept of chemical energy and the transformations of forms of energy during these processes. This problematic understanding did not seem to be limited to that student, as shown by the fact that only a small number of students (two out of 28) mentioned the relation between glucose and the production of energy for cellular work (EN3) correctly in the explanatory tasks.

Interviews with the four students conducted immediately after the posttest revealed that they were able to mention correct propositions in relation to photosynthesis and cellular respiration in a conversation when appropriate cues were offered. However, the interviewed students had problems with the concepts of biosynthesis and energy, in accordance with the results of all students as presented in Figure 5. This difficulty is illustrated by a quote from student 21. In the pretest, this student mentioned the propositions PH1, PH2, PH3, and PH4 in the defining tasks by writing down the chemical equation of photosynthesis. In addition to the four PH propositions, he also mentioned the propositions CR1, CR2, CR3, CR4, and BI2 in the defining tasks on the posttest. This might indicate that this student knew the chemical equations of photosynthesis and cellular respiration and that proteins are produced by biosynthesis. Yet in the explanatory tasks in the posttest, he only mentioned PH3 (photosynthesis produces oxygen).

Interviewer (I):	So, the question with the shrimps and algae in the sealed container. You adjusted your previous answer by adding: algae perform photosynthesis producing oxygen for the shrimps. (The student mentioned proposition PH3.)
Student (S):	Yes, I thought algae are green plants which performs photosynthesis producing oxygen needed for the shrimps to live. The first time [referring to the pretest] I hadn't thought about that.
I:	Can you tell me what is needed for photosynthesis?
S:	Water is needed, which is there anyway, and light energy, maybe from a light bulb, and carbon dioxide which must be added to the container or released from cellular respiration. (The student mentioned proposition CR4 and the first part of propositions PH1 and PH4.)
I:	So what is produced by photosynthesis that allows the shrimps to live?
S:	Oxygen and glucose. (This refers to propositions PH2 and PH3.)
I:	And can you be more specific about this light energy, where is it going when it enters the container?
S:	That is still difficult for me, because it has something to do with ATP. I think it is important for the biosynthesis which produces proteins (he mentioned BI2). But I'm quite sure the light energy is used somehow.

From this fragment, it becomes clear that this student learned to shift his locus of explanation to the (sub)cellular level of organization because he was now actively reasoning, as shown by his mentioning cellular processes. This reasoning process, however, was not observed during

the interview conducted after the pretest. Although he did seem to partially understand the relation between photosynthesis and cellular respiration, he did not understand the relation between photosynthesis and biosynthesis. This disconnect might be attributed to a misunderstanding of the conversion of energy from one form to the other. Moreover, he did not grasp the idea that molecules can contain energy. The low numbers of students who mentioned energy-related propositions (EN1-4) as shown in Figure 5 indicates that he was not alone.

The results of the explanatory questions in the final test showed high numbers of students who mentioned the following propositions correctly compared with the explanatory tasks in the posttest: PH1 (from 12 to 22), PH2 (from 2 to 24), CR1 (from 5 to 11), and CR4 (from 4 to 13). Although it is possible that the formulation of questions in the final test prompted students rather better to mention these propositions, many students appeared to make considerable progress compared with the results of the posttest. Furthermore, 23 students mentioned proposition BI1 whereas only three students mentioned this proposition in the defining tasks in the posttest. The responses showed that students learned to give biological explanations by switching to metabolic processes at the cellular level of biological organization. For instance, in a context-oriented question about the growth of algae more than 20 students mentioned the concept of photosynthesis as a glucose-producing process that is dependent on the presence of light and referred to PH1 and PH2. The number of students who mentioned proposition EN2 (five) and EN3 (one) was still low, however. Although 11 students mentioned that glucose is needed for cellular respiration only one student mentioned that this process generates the energy for cellular activity (EN3). The results of the multiple-choice questions in the final test were as follows: Two students gave a correct response to the first question, which required the propositions PH4, PH2, BI1, and BI4, and 20 students gave a correct response to the second question in which the propositions PH4, PH2, CR1, and CR4 were required. Although 23 students mentioned the individual proposition "Glucose is needed for biosynthesis" (BI1; Figure 5b), apparently it was difficult for students to relate all four propositions to each other in the multiple-choice question. This trend corresponds with the results from the interviews and indicates that most students lack the ability to mention propositions that connect two (or more) metabolic processes.

Influence of Three Context-Embedded LT Activities

In this section, we first describe how students' conceptual coherence, in terms of their ability to mention certain propositions, seemed to have developed during each LT activity. Next, we try to identify how each LT activity contributed to the growth of students' conceptual coherence. In addition, we present how students perceived the usefulness of each LT activity.

LT Activity 1: Using Graphic Visualizations. Graphic visualizations of the context (Figures 2 and 3) were used to relate differences in the production chains of meat and organic protein-rich food products to carbon dioxide emission and uptake. During the reflection phase, graphic visualizations of mitochondria and chloroplasts (Figure 4) were used to indicate the location within the cell where these processes take place. It was expected that students would learn that carbon dioxide is needed for photosynthesis (proposition PH4), that photosynthesis produces glucose (proposition PH2) which can be used for cellular respiration (proposition CR3) in plant and animal cells, that cellular respiration produces carbon dioxide (proposition CR4) which in turn can be used for photosynthesis

(proposition PH4), and that cellular respiration converts chemical energy in glucose into heat (proposition EN2) and energy for cellular activity (proposition EN3).

During the interviews, all four students (numbers 11, 13, 21, and 26) could relate the release of carbon dioxide to cellular respiration (proposition CR4) and the uptake of carbon dioxide to photosynthesis (proposition PH4). They also understood that cellular respiration takes place in both plant and animal cells, and they switched between the organizational level of the organism (plant) and the (sub)cellular level. Moreover, three students (numbers 11, 21, and 26) mentioned that during photosynthesis glucose was produced (proposition PH2) and three students (numbers 11, 13, and 26) mentioned that cellular respiration also produces energy for movement (proposition EN3). Surprisingly, they connected both processes only by mentioning carbon dioxide and did not notice glucose. One student (no. 11) said: "A plant needs carbon dioxide which is directed to the chloroplasts. There photosynthesis takes place and the part of carbon dioxide that is not used in this process is the carbon dioxide that is emitted by the plant." This student also remarked: "Carbon dioxide is converted into oxygen." None of the other students seemed to understand that the carbon (C) atom of carbon dioxide is fixed in glucose.

Possibly, when using the graphic visualization of the context (Figure 2) the initial focus on carbon dioxide emission and uptake reinforced students' ideas that the processes are mainly opposite gas exchanging processes and that cellular respiration only takes place in animals and humans. These misunderstandings seem to be associated with unfamiliarity with the concept of chemical energy and the involvement of cellular respiration in transformations of forms of energy. When asked where the energy (for movement) produced by cellular respiration came from, none of the students responded correctly. Obviously, the teacher's incorrect formulations of how energy was transformed during these processes (Table 1) were not helpful. Moreover, conservation of matter appeared to be a totally unfamiliar principle to the students. They thought that matter (C atoms) could just disappear during metabolic processes. One of the students (no. 11) said: "Carbon dioxide is turned into oxygen." When the interviewer asked: "Can you explain what happens with the C of carbon dioxide?" the student responded: "I don't have a clue."

An analysis of the video recordings of this lesson revealed that when the teacher used the graphic visualization of the cells to explain the chemical equations of the metabolic processes, he wrote on the whiteboard carbon dioxide as a chemical notation (CO_2) and spelled glucose as a word without mentioning its chemical notation ($C_6H_{12}O_6$). This might have prevented students from tracing C atoms. The teacher also did not explain the concept of chemical energy in relation to glucose. He said: "Glucose is turned partially into energy like movement and heat." This oversimplification could explain students' misunderstandings about the transformations of forms of energy and matter, which persisted after the lesson sequence, as shown in the section Changes in Mentioning Propositions Before and After Lesson Sequence. Moreover, because the teacher described the forms of energy "movement and heat" as the result of cellular respiration, students intuitively tended to relate this process to (homoeothermic) animals at the organizational level of the organism. Consequently, it is not surprising that they did not link cellular respiration to plant cells. Because during the reflection phase there was no focus on the use of energy within an animal or plant cell, e.g., the movement of cellular particles during growth, students' ideas that cellular respiration only takes place in animals (and animal cells) were reinforced.

From the interviews, it became clear that students perceived that the graphic visualizations were supportive of their learning processes. When we asked them to explain how an environmental advisor deals with carbon dioxide emissions during the production chains of meat and soya, all students mentioned that they thought immediately of the graphic visualization of the production chains (Figure 2). One student (no. 21) said, "By studying the

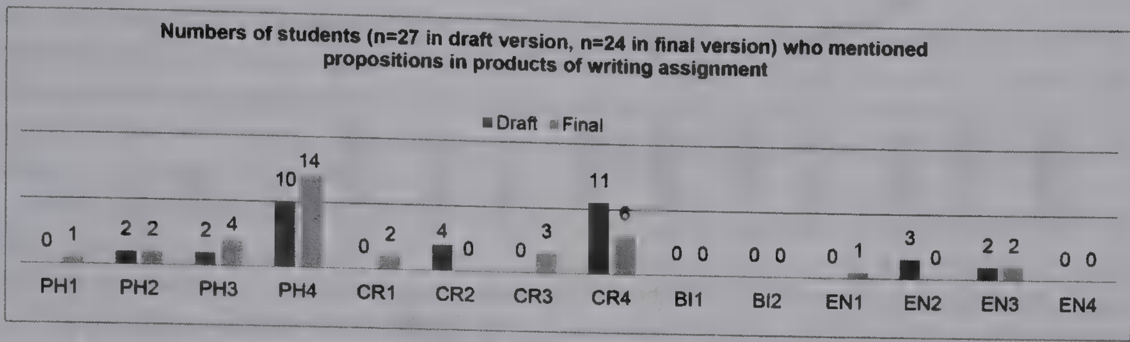


Figure 6. Numbers of students who mentioned propositions in written texts. The codes refer to the propositions shown in the reference concept map as presented in Figure 1.

graphic visualization I got an overview and by drawing the arrows it was clear to me where carbon dioxide was used and produced." The teacher also recognized this usefulness during an interview. He explained, "By showing the graphic visualization of the context I could explain in a concrete manner which activities an environmental advisor performs. Students collaborated constructively when drawing the arrows to indicate carbon dioxide uptake and release." He continued: "From carbon dioxide I could easily switch my explanation to the chemical equations of photosynthesis and cellular respiration." All students confirmed that the graphic visualizations of chloroplasts and mitochondria were very useful for locating where in a cell photosynthesis and cellular respiration take place and seeing the differences between plant and animal cells.

LT Activity 2: Writing. Students had to write a draft version (Part 1) and a final version (Part 2) of advice from the perspective of the environmental advisor. We expected that in the draft version students would mention at least propositions CR4 and PH4 (Figure 1). In the final version, we expected students to also mention propositions PH1, PH2, CR1, CR2, EN2, and EN4. From the evaluation of the steps in the research scenario (Table 1), it was evident that students asked the teacher many questions during the brainstorm phase of Part 1 of this LT activity. They seemed to experience difficulties when starting to write. Moreover, the reflection phases of both parts and Part 2 of this LT activity were not conducted as intended. In Part 1, reflection was limited to a short presentation by the teacher without interaction, whereas in Part 2 the reflection phase was not observed at all.

Figure 6 shows the numbers of students who mentioned propositions from the reference concept map. It appears that not many propositions were mentioned except for proposition PH4 (carbon dioxide is needed for photosynthesis) and CR4 (cellular respiration produces carbon dioxide). Surprisingly, propositions CR2, CR4, and EN3 were mentioned even less frequently in the final version of the written products. Furthermore, it was remarkable that some students wrote a text in which they explained the differences in carbon dioxide emissions between the production chains of animal and plant protein-rich food products without referring to cellular processes.

From the interviews conducted after Part 1 of this LT activity, it was clear that students were able to explain the aim of writing environmental advice. For example, one student (no. 21) said, "To show how the production of proteins in both plants and animals leads to carbon dioxide emissions." The interviews conducted after Part 2 of this LT activity showed that the four students were able to mention most of the individual propositions but found it hard to apply these propositions in the advice. Student 21 said, "I don't know how these cellular processes are involved in the growth of animals." Moreover, interrelating cellular respiration and biosynthesis was difficult. Students seemed to regard the conversion of

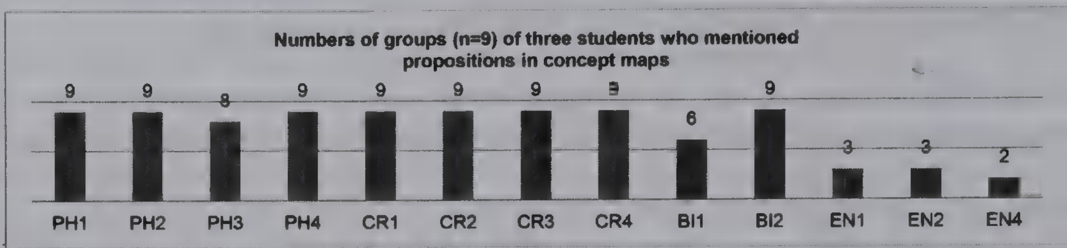


Figure 7. Numbers of groups consisting of two, three, or four students who mentioned propositions related to the concepts of photosynthesis (PH1–4), cellular respiration (CR1–4), biosynthesis (B1 and B2), and energy (EN1, EN2, and EN4) in products of a concept-mapping activity. The codes refer to the propositions shown in the reference concept map as presented in Figure 1.

glucose into either carbon dioxide or proteins as unidirectional processes. One student (no. 21) said, "As soon you have the proteins, they can't be used for cellular respiration and turned into carbon dioxide anymore." Another student (11) said, "If a chicken eats proteins, it grows because proteins are building blocks." It appears that ignorance of digestion processes, chemical structures of molecules, and other chemical breakdown processes (besides cellular respiration) hinders students in establishing a clear line of reasoning.

Although students understood the aim of the advice, they all said they needed more support at the beginning of the writing process. This mainly concerned the first part of this LT activity in which students had to write a draft. One student (11) said, "After the teacher explained which cellular processes had to be used in the advice it became clear to me. I found it hard to find the appropriate information from the manual myself." Apparently, more guidance was needed to switch from the context to the (underlying) concepts. When asked how they valued both writing activities, the four students agreed that they considered them useful. Student no. 26 explained, "The writing activities forced me to think about everything, so I learned from it. The written text was a kind of summary of the lesson sequence." The postlesson evaluations revealed, however, that the students had divergent opinions about the usefulness of the writing assignments. Ten students mentioned that the assignments were too difficult for them. This was also confirmed by the teacher who said,

> The step from constructing chemical equations during my explanation to integrating these equations in a text is too big for many of these students. They had to explain differences in energy loss and carbon dioxide release during protein production by connecting these complex chemical processes. During the brainstorm phase I noticed this was very difficult for them and, although they really tried, most of them did not know how to start writing their texts. I think the time spent on this topic in advance of the writing activity was too short.

LT Activity 3: Concept Mapping. In this LT activity, nine groups of students constructed two concept maps over the course of two lessons. The focus questions of these concept maps were as follows: *How do plant cells produce proteins?* and *How do animal cells produce proteins?* We expected that students would show the following propositions in their concept maps: PH1 to PH4, CR1 to CR4, BI1 and BI2, and EN1, EN2 and EN4 (Figure 1). This LT activity was broken into several steps spanning the mapping process. In the reflection phase, it was intended that the students would be prompted to explain the differences between the processes in plant cells and animal cells. Only this last step was not recognized as intended (Table 1).

Figure 7 presents the groups of students who showed propositions in their concept maps. It was mainly the energy propositions (EN1, EN2, and EN4) which were incorrect or

lacking. Obviously, students' conceptual understanding of the core concept of energy was still problematic. The teacher recognized this trouble, saying in an interview, "I noticed that many students had problems with the term chemical energy and even more with ATP." It also appeared that the concept maps helped the teacher to point out exactly which propositions were problematic for students.

In interviews conducted after this LT activity, students were asked to explain how the three metabolic processes are involved in protein production in plant and animal cells. All four students explained the relation between photosynthesis and cellular respiration. They seemed to feel confident when talking about these processes, but they all had problems when linking photosynthesis to biosynthesis. The role of glucose as a "building block" to produce proteins was not clear to them. Student no. 26 said: "I think the proteins in animal cells are there because animals consume protein-rich food." He also remarked that biosynthesis was still a rather vague concept. Student no. 21 thought that only ATP and minerals are needed for biosynthesis to produce proteins. When asked to explain why photosynthesis is the basis for protein production only student no. 11 was able to connect this to the production of glucose needed for biosynthesis.

Students rated the concept-mapping activity as very useful, and this rating was confirmed by almost all students in the postlesson evaluations. Students specifically highlighted interacting in groups and comparing and discussing their own concept maps with those constructed by other groups as positive aspects of this LT activity. Moreover, during the mapping activity the teacher prompted students to explain the relations between concepts and gave feedback.

DISCUSSION

The two central research questions addressed in this study were as follows: *How does students' conceptual coherence develop for students during a context-based lesson sequence?* and *How do context-embedded LT activities influence the development of conceptual coherence?* Following a design-based research approach, we described how a context-based lesson sequence in biology was designed, conducted, and evaluated on its practicability and effectiveness. Next, we discuss reported changes in mentioning propositions (as presented in the section Changes in Mentioning Propositions Before and After Lesson Sequence) and give explanations for unexpected findings. Then, we reflect on the three examined context-embedded LT activities (as presented in the section Influence of Three Context-Embedded LT Activities). Finally, we reflect upon the usefulness and limitations of the research scenario and the reference concept map with a view to future research on context-based science education.

Development of Conceptual Coherence

The significant gains in test scores for mentioning propositions from the reference concept map indicated that students' conceptual coherence developed during the lesson sequence. More specifically, students were better able to mention propositions related to just one of the core concepts than propositions between core concepts. Mentioning propositions that included the core concept of energy appeared extremely difficult, even at the end of the lesson sequence. Other consistent findings in our data included students' attempts to mention more concepts and propositions in their explanations. This indicates that they improved their abilities to switch their reasoning to the cellular level of biological organization.

Nevertheless, the results indicated limited gains in mentioning propositions. This is slightly surprising because much effort was made during the design process to support

students in overcoming the reported learning problems on this topic. It is possible that the conceptual framework as presented in the reference concept map might have been too difficult for many of these students at this level of education, especially given in the short time span of ten lessons. In a study on similar explanatory tasks, Songer and Mintzes (1994) showed that even between novice and experienced university biology students, there were hardly any significant increases between their ideas on photosynthesis and cellular respiration. Difficulties in both teaching and learning this topic were also recognized by Mohan et al. (2009) who reported that even by the end of high school no more than half of the students were attempting to use, more or less consistently, chemical processes to explain macroscopic and large-scale events. Only 10% of the students distinguished matter from energy during metabolic processes. The other 90% of the students were not able to describe chemical changes based on scientific principles such as the conservation of matter and energy. These results are in line with our findings.

We found that alternative (incorrect) understandings concerning the energy concept did not disappear. Students still had problems interrelating two (or more) metabolic processes, and they did not understand the conservation and transformations of matter and forms of energy in metabolic processes. Lin and Hu (2003), who showed that 13-year-old students failed to interrelate biological concepts concerning energy flow and matter cycling, pointed out that a lack of chemical and physical interpretations when teaching biology is one of the causes of this failure. Moreover, it seemed that at the beginning of the lesson sequence there was a lack of basic understanding of the concepts of energy and matter. Therefore, we recommend that lessons in chemistry, physics, and biology pay more attention to providing a common base for these concepts. For instance, in 9th-grade science lessons the focus could be on conversions of forms of energy and matter in meaningful contexts. In addition, we recommend that the conceptual network (or parts of it) be frequently revised during the course of the curriculum. There are opportunities to relate the conceptual network to a variety of context-areas such as health, sport, and environment.

As an explanation for the limited gains in mentioning particular propositions (and thus in the development of conceptual coherence), we think that the order of the contexts in the lesson sequence restricted the order in which the concepts and propositions from the reference concept map were introduced. This is illustrated with the following two examples.

First, in the beginning of the lesson sequence too much attention was given to the concepts of carbon dioxide and oxygen. This emphasis supported students' intuitive ideas. They persisted in thinking that photosynthesis and cellular respiration are solely opposite gas-exchanging processes (Cañal, 1999), without regarding them as energy-transforming processes. In the beginning of the lesson sequence, there was a lack of attention to the chemical notation of glucose ($C_6H_{12}O_6$) and the idea that the C-bindings contain energy in a chemical form. Because of this oversight, students found it hard to trace matter (C atoms) and energy once the three metabolic processes were introduced.

We therefore propose that the beginning of the lesson sequence should focus more intensively on glucose as a connecting concept between the metabolic processes. Glucose can be regarded as a *threshold concept* because it "opens up a new and previously inaccessible way of thinking about something" (Roseman, Stern, & Koppal, 2010). We expect that if students understand the chemical structure of glucose, this understanding will support them in learning the concepts of chemical energy and transformations of forms of energy and matter in relation to the three metabolic processes. One possibility is to extend the first context with an LT activity in which students build a chemical model of glucose (paying attention to the energetic bindings between the C atoms) to find out why humans need organic substances for energy. Such LT activity could be conducted after the role-play in which is stated that people need meat or meat substitutes for energy.

Second, the introduction of the concepts of biosynthesis and chemical energy lasted until the end of the second context. Therefore, relatively less time was available to master these concepts (compared with photosynthesis and cellular respiration). This suggests that the number of concepts that are introduced in a lesson sequence should be balanced with the number of lessons. According to the theoretical basis of the concept-context approach, however, it is preferable that there is a logical reason to introduce a new concept (Boersma et al., 2007). This dilemma is bound up inextricably with the design and study of a single context-based lesson sequence as presented in this paper. A sequence of contexts in a spiral curriculum would solve this problem. Then each concept could be drip-fed into the lesson and revisited in-depth and in more than one context.

Influence of Context-Embedded LT Activities on Development of Conceptual Coherence

Our evaluation of the first LT activity *(using graphic visualizations)* showed convincingly that graphic visualizations help students to structure their thoughts and to link contexts to concepts. They were used for a hands-on activity (Figures 2 and 3) and for a classroom discussion (Figures 2–4). However, the graphic visualizations did not reveal conceptual misunderstandings with respect to transformations of forms of energy and matter. Possibly, a third graphic visualization in which glucose and other substances at the molecular level are depicted would be helpful.

From the evaluation of the second LT activity (*writing*), it became clear that establishing propositions when reasoning during an individual writing process proves to be difficult for many students. Although students understood the context and the goals of the advice that had to be written, most of them were not able to use biological concepts and propositions in their writings. Moreover, we observed much variation between students in their abilities to write a text. This variation could be associated with metacognitive abilities required to structure writing. Therefore, differentiation and scaffolding is needed, for instance, by providing sample sentences to individual students on demand.

During the third LT activity (*concept mapping*), students showed the intended propositions, with the exception of propositions related to the core concept of energy. It is not surprising that students showed many propositions because they constructed the concept map during a group discussion and all concepts had already been given. This result is in line with previous research showing that concept mapping is a supportive LT activity that promotes active thinking and construction of new propositions (Nesbit & Adesope, 2006). Moreover, having students express conceptual thinking allowed the teacher to ask questions and to provide feedback. However, the teacher did not respond adequately to propositions that were difficult for students. In future implementations, it might be useful to offer the teacher a set of feedback questions to help students check their concept maps and focus their attention on the way they established propositions, including energy-related concepts.

From the evaluated research scenario, it became clear that none of the reflection phases of these three LT activities were conducted as intended. More guidance on structuring reflection is apparently needed. A well-thought questioning strategy could be helpful here. Such a strategy should elicit what students think, encourage them to elaborate on their previous answers and ideas, and help them to construct conceptual knowledge. The teacher should be prepared in terms of which questions to ask and the sort of responses to be expected from students. We suggest that parts of the reference concept map underlying the contexts can function as a "roadmap" to structure such a question strategy and help teachers to adapt questions to students' answers. The "question-based discourse" analytic framework developed by Chin (2006) would be useful to stimulate productive thinking

during reflective moments. Therefore, as a specification for the fourth design principle (reflection on concept), we advise the use of a questioning strategy when reflecting on conceptual relationships in a context.

Reflection on Use of the Reference Concept Map and the Research Scenario

Studying the influence of a context-based learning environment on conceptual learning is a challenge, which is also recognized in other context-based projects (e.g., Pilot & Bulte, 2006). We showed that a design-based research approach can shed light on the mechanisms involved in teaching strategies and learning processes. This section reflects on the usefulness and limitations of two innovative elements in our design-based research approach: the reference concept map (Figure 1) and the research scenario (Table 1).

Reference Concept Map. The reference concept map was used in two ways: to guide the design process and to assess students' learning outcomes. Because the reference concept map was the result of a systematic analysis of school books, literature, and discussions with experts, it functioned as a theoretically and empirically underpinned framework from which learning objectives could be derived. During the design process, decisions could be legitimized by pointing out which concepts and propositions from the reference concept map were involved. Therefore, implicit decisions about the selection of social practices, the transformation of these social practices into contexts suitable for integration into the lesson sequence, and the structuring of promising LT activities could be explicated.

With respect to the evaluation of students' learning outcomes, this study showed that using the reference concept map as an assessment tool gave a clear focus for analysis. Multiple types of data sources, including those which were derived from a naturalistic setting, were analyzed on the occurrence of (intended) propositions systematically and with high validity, as indicated by the high Cohen's kappa values (see the section Research Scenario and Evaluation). Analyzing these multiple data sources in a unified way allowed triangulation. Although analyzing changes in concept maps of individual students could be a useful—and often logical—methodological approach to measuring conceptual coherence (Pearsall, Skipper, & Mintzes, 1997; Novak, 2005), this would not have been suitable for the purposes of the study presented here. Testing the effects of concept-mapping activities would have clouded the learning effects of the intervention, all the more because the reported intervention covered a relatively short period of only 10 lessons.

There were two limitations in the way; we used the reference concept map to assess students' conceptual understanding, however. First, different types of data sources cannot be compared one to one with high validity if they are not identical. For instance, the extreme rise of PH2 in the final test compared with the explanatory tasks in the posttest (from 3 to 24, Figure 5b) could suggest that the cues that prompted students to mention this proposition were different. In the posttest, this proposition was probably elicited more easily. Second, scoring only correct propositions from the reference concept map seems to give only a rough indication of students' conceptual coherence. A subtle improvement in students' understanding, such as the observation that students gradually shifted their locus of explanation from the organism to the (sub)cellular level of biological organization, could not be detected. We observed that students often used common language instead of biological terms ("Animals need food to burn and to get energy"). On the other hand, when students mentioned the biological terms as presented in the reference concept map they often did not formulate the exact proposition ("Sunlight is needed to produce glucose").

These indications of a certain development of conceptual coherence were not detected with our reference concept map–based assessment method. This might also explain the relatively low number of students who showed improvement in the posttest and, to a lesser extent, in the final test.

For future research, an even more discriminating assessment tool could be developed in which each proposition is subdivided into progressive levels of understanding. Students' responses could be categorized into these levels. For instance, when a student remarks, "A plant needs sunlight to produce food" this statement is correct but does not prove an understanding of the involvement of metabolic processes at the cellular level of biological organization. Therefore, this remark could be seen as an intermediate level of understanding of the propositions PH1 (energy from sunlight is needed for photosynthesis) and PH2 (photosynthesis produces glucose). This would require test items to be developed carefully to evoke responses in which different levels of propositional understanding could be recognized.

Moreover, different test items that trigger students to mention the same proposition should be validated. Furthermore, analyzing the degree to which students mention combinations of propositions in one response would indicate their conceptual coherence at a higher level. We should, however, be aware that a single assessment tool always gives a limited interpretation of student's understanding (Mintzes et al., 2005). We have shown in this paper that a reference concept map is a robust instrument to analyze multiple data sources obtained through multiple means of data collection, resulting in a coherent description of the learning and teaching processes.

Research Scenario. The aim of the research scenario was twofold. First, because it was constructed in parallel with the design of the lesson sequence the designers were forced to carefully consider each step of the lesson sequence. Second, it was used to evaluate the design on its practicability (Ummels et al., 2015). This evaluation (section Research Scenario and Evaluation and Table 1) provided insight into the quality of each of the LT activities within the lesson sequence. Steps that were conducted as intended by the teacher but did not result in the intended learning outcomes gave useful input for adapting the written design. If steps were not conducted as intended by teacher or students, the learning outcomes had to be considered carefully.

This analysis made it clear that the specific actions of the teacher are extremely important to students' conceptual learning. For instance, the language and phrasing used by the teacher when explaining complex concepts like the energy concept need to be balanced: On the one hand, simplifications of a complex topic can all too easily confirm students' misconceptions; on the other hand, students become lost when the teacher gives too much content-specific information. In conclusion, for future design-based research on learning concepts within context-based education, we recommend both the use of a reference concept map and a research scenario to keep a focus on the research aims and to gain an in-depth understanding of teaching strategies and learning processes.

The authors acknowledge the following people's contributions to this article: Sjef van Nilwik for conducting the reported lesson sequence with such commitment, Eddie Denessen for his support with the statistical analysis and helpful comments on the structure of this article, Quincy Elvira for his critical remarks on this article, and Horst Wolter for the creative illustrations presented in Figures 2–4, which were constructed in close cooperation with the authors.

REFERENCES

Amir, R., & Tamir, P. (1990). Detailed analysis of misconceptions as a basic for developing remedial instruction: The case of photosynthesis. Paper presented at the Annual Meeting of the American Educational Research Association, April 1990, Boston, MA.

Ausubel, D. P. (1968). Educational psychology. A cognitive view. New York: Holt, Rinehart & Winston.

Barker, V., & Millar, R. (2000). Students' reasoning about basic chemical thermodynamics and chemical bonding: What changes occur during a context-based post-16 chemistry course? International Journal of Science Education, 22(11), 1171–1200.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. Science Education, 91(3), 347–370.

Boersma, K. T., van Graft, M., Harteveld, A., de Hullu, E., de Knecht-van Eekelen, A., Mazereeuw, M., van den Oever, L., & vander Zande, P. A. M. (2007). Leerlijn biologie van 4 tot 18 jaar. Uitwerking van de concept-contextbenadering tot doelstellingen voor het biologieonderwijs [Biology curriculum for ages 4 to 18. Elaboration of the concept-context approach to learning goals for biology education]. Utrecht, The Netherlands: NIBI.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000a). How experts differ from novices. In J. D. Bransford, A. L. Brown, &, R. R. Cocking (Eds.), How people learn (pp. 31–50). Washington, DC: National Research Council.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000b). Learning and transfer. In J. D. Bransford, A. L. Brown, &, R. R. Cocking (Eds.), How people learn (pp. 51–78). Washington D.C.: National Research Council.

Brown, M. H., & Schwartz, R. S. (2009). Connecting photosynthesis and cellular respiration: Preservice teachers' conceptions. Journal of Research in Science Teaching, 46(7), 791–812.

Cañal, P. (1999). Photosynthesis and "inverse respiration" in plants: An inevitable misconception? International Journal of Science Education, 21(4), 363–371.

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. International Journal of Science Education, 28(11), 1315–1346.

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. Journal of Research in Science Teaching, 44(6), 815–843.

CvE. (2009). Biologie HAVO Syllabus centraal examen 2011 [Biology general secondary education syllabus national exam 2011]. Retrieved from https://www.examenblad.nl/examenstof/syllabus-2011-biologie-havo/2011/f=/syllabus_biologie_havo_2011_20091002.pdf. Acces date: 10 May 2015.

Davidowitz, B., & Rollnick, M. (2001). Effectiveness of flow diagrams as a strategy for learning in laboratories. Australian Journal of Education in Chemistry, 57, 18–24.

DiSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. Cognitive Science, 28(6), 843–900.

Fisher, K. M. (2001). Meaningful and mindful learning. In K. M. Fisher, J. H. Wandersee, & D. E. Moody (Eds.), Mapping biology knowledge (pp. 77–94). Dordrecht, The Netherlands: Kluwer.

Flores, F., Tovar, M. E., & Gallegos, L. (2003). Representation of the cell and its processes in high school students: An integrated view. International Journal of Science Education, 25(2), 269–286.

Galbraith, D. (1999). Writing as a knowledge-constituting process. Amsterdam: Amsterdam University Press.

Gilbert, J. K. (2006). On the nature of "context" in chemical education. International Journal of Science Education, 28(9), 957–976.

Gilbert, J. K., Bulte, A. M. W., & Pilot, A. (2011). Concept development and transfer in context-based science education. International Journal of Science Education, 33(6), 817–837.

Keselman, A., Kaufman, D. R., Kramer, S., & Patel, V. L. (2007). Fostering conceptual change and critical reasoning about HIV and AIDS. Journal of Research in Science Teaching, 44(6), 844–863.

Kinchin, I. M. (2011). Visualising knowledge structures in biology: Discipline, curriculum and student understanding. Journal of Biological Education, 45(4), 183–189.

King, D., & Ritchie, S. M. (2012). Learning science through real-world contexts. In B. J. Fraser, K. G. Tobin, &, C. J. McRobbie (Eds.), Second international handbook of science education (pp. 69–79). Dordrecht, The Netherlands: Springer.

Klaassen, C. (1995). A problem-posing approach to teaching the topic of radioactivity. Doctoral dissertation, Utrecht University. Utrecht, The Netherlands: Utrecht University.

Kose, S., Usak, M., & Bahar, M. (2009). A cross-age study of students' understanding and their misconceptions about plant nutrition. Didactica Slovenica-Pedagoska Obzorja, 24(1), 109–122.

Lijnse, P., & Klaassen, C. (2004). Didactical structures as an outcome of research on teaching-learning sequences? International Journal of Science Education, 26(5), 537–554.

Lin, C., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. International Journal of Science Education, 25(12), 1529–1544.

McCabe, B. (2011). An integrated approach to the use of complementary visual learning tools in an undergraduate microbiology class. Journal of Biological Education, 45(4), 236–243.
McKenney, S., & Reeves, T. C. (2012). Conducting educational design research. London: Routledge.
McMichael, A. J., Powles, J. W., Butler, C. D., & Uauy, R. (2007). Energy and health 5—Food, livestock production, energy, climate change, and health. Lancet, 370(9594), 1253–1263.
Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). Assessing science understanding. London: Elsevier Academic Press.
Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. Journal of Research in Science Teaching, 46(6), 675–698.
Nesbit, J., & Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. Review of Educational Research, 76(3), 413–448.
Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. Research in Science Education, 35(1), 23–40.
Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct them. Pensacola, FL: Institute for Human and Machine Cognition.
Novak, J. D., Mintzes, J. J., & Wandersee, J. H. (2005). Learning, teaching and assessment: A human constructivist perspective. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), Assessing science understanding (pp. 1–13). London: Elsevier.
Ogborn, J. (1997). Constructivist metaphors in science learning. Science and Education, 6(1–2), 121–133.
Pearsall, N. R., Skipper, J. E. J., & Mintzes, J. J. (1997). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. Science Education, 81(2), 193–215.
Pilot, A., & Bulte, A. M. W. (2006). The use of "contexts" as a challenge for the chemistry curriculum: Its successes and the need for further development and understanding. International Journal of Science Education, 28(9), 1087–1112.
Ritchhart, R., Turner, T., & Hadar, L. (2009). Uncovering students' thinking about thinking using concept maps. Metacognition Learning, 4, 145–159.
Roseman, J. E., Linn, M. C., & Koppal, M. (2008). Characterizing curriculum coherence. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), Designing coherent science education (pp. 13–36). New York: Teachers College, Colombia University.
Roseman, J. E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of high school biology textbooks. Journal of Research in Science Teaching, 47(1), 47–70.
Songer, C. J., & Mintzes, J. J. (1994). Understanding cellular respiration—An analysis of conceptual change in college biology. Journal of Research in Science Teaching, 31(6), 621–637.
Southerland, S. A., Smith, M. U., & Cummins, C. L. (2005). "What do you mean by that?" Using structured interviews to assess science understanding. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), Assessing science understanding (pp. 71–93). London: Elsevier.
Tsai, C. C. (2000). The effects of STS-oriented instruction on female tenth graders' cognitive structure outcomes and the role of student scientific epistemological beliefs. International Journal of Science Education, 22(10), 1099–1115.
Tytler, R. (2005). School innovation in science: Change, culture, complexity. In K. Boersma, M. Goedhart, O. De Jong, & H. Eijkelhof (Eds.), Research and the quality of science education (pp. 89–105). Dordrecht, The Netherlands: Springer.
Ummels, M., Kamp, M., de Kroon, H., & Boersma, K. T. (2013). De ontwikkeling van conceptuele samenhang binnen concept-contextonderwijs. Een case study voor het vak biologie in 4-havo [Developing conceptual coherence within concept-context-based education. A case study in senior general secondary biology education]. Pedagogische Studiën, 90, 19–32.
Ummels, M., Kamp, M., de Kroon, H., & Boersma, K. T. (2015). Designing and evaluating a context-based lesson sequence promoting conceptual coherence in biology. Journal of Biological Education, 49(1), 38–52.
Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). Educational design research. London: Routledge.
Van Oers, B. (1998). From context to contextualizing. Learning and Instruction, 8(6), 473–488.
Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), The collected work of L. S. Vygotsky (pp. 39–285). New York: Plenum Press.
Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In G. L. Gabel (Ed.), Handbook of research on science teaching and learning (pp. 177–210). New York: Macmillan.
Wierdsma, M. (2012). Recontextualising cellular respiration. Doctoral dissertation, Utrecht University. Utrecht, The Netherlands: Utrecht University.

Science
Education

The Use of Modeling-Based Text to Improve Students' Modeling Competencies

JING-PING JONG,[1,2] MEI-HUNG CHIU,[1] SHIAO-LAN CHUNG[1,3]
[1]*Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan 117;* [2]*New Taipei Municipal Jinhe High School, Zhonghe District, New Taipei City, Taiwan 235;* [3]*New Taipei Municipal New Taipei Senior High School, Sanchong District, New Taipei City, Taiwan 241*

Received 17 December 2013; accepted 16 February 2015
DOI 10.1002/sce.21164
Published online 14 May 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This study investigated the effects of a modeling-based text on 10th graders' modeling competencies. Fifteen 10th graders read a researcher-developed modeling-based science text on the ideal gas law that included explicit descriptions and representations of modeling processes (i.e., model selection, model construction, model validation, model analysis, model deployment, and model reconstruction) and submicroscopic perspectives of gas particles. The results revealed that the students not only developed their modeling competencies but also constructed scientific mental models of the ideal gas law after reading the modeling-based text. On the basis of their mental models, students interpreted macroscopic phenomena with submicroscopic concepts of gas particles for some modeling stages. This study demonstrates that modeling-based text enables students to better apply scientific information in the construction of their conceptual knowledge and helps them develop their modeling competencies. © 2015 Wiley Periodicals, Inc. *Sci Ed* **99:**986–1018, 2015

Correspondence to: Mei-Hung Chiu, e-mail: mhchiu@ntnu.edu.tw
Contract grant sponsor: Ministry of Science and Technology in Taiwan.
Contract grant numbers: 99-2511-S-003-024-MY3 and 102-2511-S-003-006-MY3.

INTRODUCTION

Scientists construct scientific models to explain natural and physical phenomena and to formulate scientific theories. They work with various types of models to search for relationships among parameters to find the patterns or mechanisms underlying specific phenomena (Giere, 1988; Halloun, 2004). Given the high value placed on the use of models in science, it is important to teach students about how scientists use models in scientific research and to develop students' own competencies for using models in science learning. The literature shows that students often possess incorrect mental models and experience difficulty applying scientific models to novel and even similar contexts (Chang & Chiu, 2009; Clement, 1989, 2008; Dori & Kaberman, 2012; Grosslight, Unger, Jay, & Smith, 1991; Hestenes, 1992; Schwarz et al., 2009; Schwarz & White, 2005). Currently, there are few formal teaching activities (e.g., integrating the history of science, direct teaching, and outdoor inquiry) that explicitly target the epistemological views of scientific models and the development of modeling competencies to improve students' mental modeling (Chang & Chiu, 2009; Gobert & Pallant, 2004; Hestenes, 1992; Justi & Gilbert, 2002; Saari & Viiri, 2003; Schwarz et al., 2009; Schwarz & White, 2005). Through such activities, students directly experience and come to understand how scientists generate, test, validate, and modify scientific models. However, the research on students' modeling competencies has not examined whether particular stages in the modeling process need to be emphasized over others to maximize student learning. In addition, the effects of the use of modeling-based texts for students' science learning have not been explored.

THE NEED FOR A MODELING-BASED TEXT

In many countries, scientific texts are the most popular sources of pedagogical materials in school settings. However, research shows that scientific texts inadequately consider the multidimensionality of scientific expertise such as the nature of science, the process of scientific modeling, and conceptual incoherence of the use of multiple history models (Gericke & Hagberg, 2010; Gericke, Hagberg, & Jorde, 2013; Smith & Adkison, 2010). Many studies concerning scientific texts support the use of conceptual change, such as conceptual change or refutational texts for fostering learning and removing alternative conceptions (Chambers & Andre, 1997). A conceptual change text asks students to predict and explain what will happen in a given situation. Such texts ideally make students aware of their misconceptions and cause them to reconstruct their mental models. Refutational texts pinpoint students' misconceptions and provide the correct explanation for the presented phenomenon (Palmer, 2003; Tippett, 2010). Studies on conceptual change and refutational texts stress students' individual misconceptions rather than system-level concepts regarding scientific expertise and mental modeling. However, it is necessary for students to understand how scientists recognize and use systems thinking in scientific enterprises. This study extends the research on texts in science teaching and learning by investigating the impact of a modeling-based text on 10th graders' modeling competencies.

This literature review includes three areas: (a) modeling in science education, (b) evaluation of modeling competencies, and (c) research on the ideal gas law. Based on the literature, we designed a modeling-based text and analyzed students' modeling competencies after exposure to the modeling-based text.

Modeling in Science Education

Modeling is the process whereby scientists and students generate, construct, revise, and reconstruct mental models that allow them to solve problems and conceptualize scientific

knowledge. While this general sequence is supported by the literature, different researchers use slightly different terminology to describe its stages (Clement, 1989, 2000; Halloun, 1996, 2004; Hestenes, 1995, 2010; Schwarz et al., 2009). For example, Clement (1989, 2000) stated that the modeling sequence included model generation, evaluation, and modification. Regardless of the differences in terminology, there is consensus around the sequence being recursive until it results in the construction of a useful internal model.

Hestenes (1995, 2010) and Halloun (1996, 2004) described the sequence of modeling as including model selection, construction, validation, analysis, and deployment. That is, the first stage in modeling for students consists of identifying and describing the components of the phenomenon of interest. In the second stage, students are guided to construct a specific model that helps them solve the problem at hand. Then, students assess the internal and external consistency of the model they just constructed. Once a model is validated, students collect and analyze data to answer the problem for which the model was designed. Then, students transfer their model to a similar or novel situation. It is believed that this focus on developing modeling competencies in turn helps students learn about the core ideas of science. However, what is currently unknown is whether any particular stage(s) in the modeling sequence requires greater focus on the part of students in order for students to understand and effectively use modeling in their learning of scientific concepts.

Evaluation of Modeling Competencies

The term modeling competencies is widely used, despite the lack of a single clear definition. Gilbert (2005) stated that modeling competencies of metavisual representation mainly acquire, retain, retrieve, amend, and monitor the corresponding process and outcomes. Other researchers recognize that modeling competencies, such as generating, revising, and evaluating, are necessary for models to develop from naive levels of sophistication to scientific models (Chang & Chiu, 2009; Dori & Kaberman, 2012; Kaberman & Dori, 2009; Wang & Barrow, 2011; Zöttl, Ufer, & Reiss, 2011). Kaberman and Dori (2009) describe modeling competencies as thinking skills that not only generate correct three-dimensional representations of the spatial structures of molecules but also transfer between different molecular representations. Wang and Barrow (2011) defined five characteristics of modeling competencies: (a) generating a mental model; (b) reconstructing, manipulating, or adjusting a generated mental model; (c) analyzing a problem and recognizing conditions and propositions; (d) monitoring the reasoning process; and (e) self-checking using an alternative approach. In another study, Schwarz and her colleagues (2009) integrated metamodeling knowledge and elements of the practice into students' practice of modeling. Students were expected to construct, use, evaluate, and revise models. In sum, modeling competencies refer to the thinking skills that students use to generate, validate, revise, and reconstruct their mental models. The main difference between good and poor modelers is the degree of metacognitve competency reflected in the construction and components of their models.

Two methods (interviews and written tests) were utilized to detect students' modeling competencies. Generally, students perform less well on interviews than on written tests. This is because interviews require students to organize their thoughts whereas written tests involve more facile recognition of presented test items. The current literature relies heavily on the use of interviews to investigate students' modeling competencies. Interviews, while considered time consuming, do have the benefit of providing in-depth information about students' understanding and preventing an overestimation of student knowledge (e.g., Chang & Chiu, 2009, Dori & Kaberman, 2012; Wang & Barrow, 2011). As such, interviews were used in this study because paper and pencil tests were not sufficient to uncover how

students used the modeling stages presented in the modeling-based text to construct their understanding of the ideal gas law. Dori and Kaberman (2012) examined students performing a transformation between chemical representations that involved mastery of four levels of chemistry understanding: (a) drawing and transferring between a molecular formula, a structural formula, and a model; (b) simple and complex molecules to model drawing; (c) symbols to submicroscopic and macroscopic (referring to observable or measurable) levels; and (d) symbols to the process level. Wang and Barrow (2011) scored five characteristics of modeling competencies using a 2, 1, or 0 based on the quality of each characteristic as revealed during students' interviews. A rubric scheme of modeling competencies provides an opportunity to look in-depth at the process of students' construction of a correct mental model; however, learning a model goes beyond learning representations (Kennedy, 2012; Knuuttila, 2011) to include the relationship between subcomponents. Instead of isolating characteristics of students' performances, it is worthwhile to confirm students' mental models and modeling competencies based on the relationship between a model's subcomponents.

The Ideal Gas Law

The ideal gas law consists of the ideal gas model and the ideal gas equation ($PV = nRT$). The ideal gas equation is itself derived from multiple scientific laws, including Boyle's law, Charles and Gay-Lussac's law, and Avogadro's law. To understand the complexities of the ideal gas equation, the student has to understand the nature and behavior of gas particles and be able to apply these principles to other concepts to understand more complex scientific concepts, such as vapor pressure and chemical equilibrium. The literature shows students' major misconceptions about the ideal gas law include misusing the ideal gas law and not being able to understand the appropriate behavior of gas particles (Kautz, Heron, Loverude, & McDermott, 2005a, 2005b; Lin, Cheng, & Lawrenz, 2000). However, most teaching of the ideal gas law focuses on the equations and formulas and ignores the nature and behavior of gas particles (de Berg, 1989). It is suggested that in-depth conceptual understanding involves making connections between chemical principles and the various chemical representations that describe phenomena. If students understand the nature of the ideal gas model and the mathematical equation of the ideal gas law, then students will be able to use this understanding to calculate and also describe what happens to gas particles when parameters are changed (Niaz, 1995; Niaz & Robinson, 1992). Kautz et al. (2005a, 2005b) provided problem-based activities to facilitate students' formation of appropriate explanations of the ideal gas law at the macroscopic and submicroscopic levels. Similarly, Levy and Wilensky (2009) used agent-based modeling designs to offer students an opportunity to construct relationships between microscopic representations and macroscopic phenomena and to interact with various models (such as conceptual models and mathematical models) that corresponded to the physical world. However, the field lacks texts to make connections between the contexts of independent activities. As a result, incorporation of systemic perspectives of science learning, such as the modeling-based approach, into chemistry teaching materials, including textbooks, is needed.

RATIONALE AND CURRENT STUDY

The present study is the first to explore the effects of a modeling-based text on students' modeling competencies regarding the ideal gas law. Although there is promising evidence that modeling-based instructional activity can help students improve their modeling competencies related to the stages of constructing, refining, and reconstructing their scientific

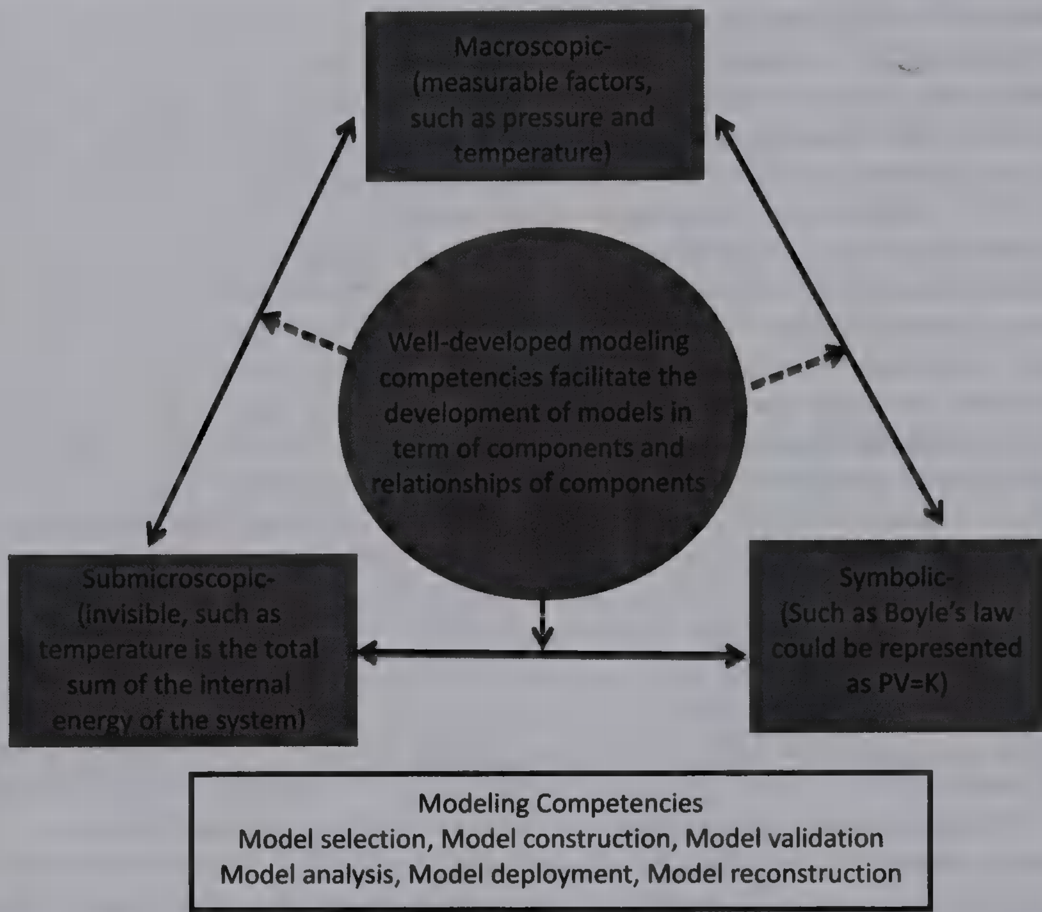


Figure 1. The conceptual framework of the study.

models, few studies stress reconstruction or transformation from model A to model B. This suggests that more research is needed to discern which stages in the modeling process require the most attention during instruction.

On the basis of argumentation, scientists revise their false mental models into correct scientific models. When facing conflicting or inconsistent results, scientists have to revise or abandon their mental models and then revise, replace, or even reconstruct their scientific models. Therefore, we argue that the way scientists reconstruct scientific models can be a channel for teaching students about scientific enterprise. On the basis of this understanding, we designed a modeling-based text that emphasizes and explicitly presents the modeling sequence and its stages, including a specific focus on model validation and model reconstruction. In practice, we adapted the modeling process (Liu & Chiu, 2010), which was modified from the modeling process proposed by Hestenes (1995) and Halloun (1996, 2004), to design the modeling-based texts that explicitly provided the stages of the modeling process, including model selection (MS), model construction (MC), model validation (MV), model analysis (MA), model deployment (MD), and model reconstruction (MR).

One of the main goals of chemistry teaching is to facilitate students' understanding of the submicroscopic representations used to explain natural phenomena. We took this into consideration when we designed the modeling-based texts on the ideal gas law (see Figure 1). The arrows between the macroscopic, symbolic, and submicroscopic elements convey the development of chemical knowledge. The arrows among the models, modeling competencies, and triple perspectives show the pathways to construct the components and the relationships of chemical knowledge on the basis of modeling processes and the

corresponding modeling competencies. We expected that students exposed to the modeling-based text would not only construct and revise their mental models but also explain macroscopic phenomena on the basis of submicroscopic representations. On the basis of the six modeling stages outlined above, we rewrote the ideal gas law text from a high school chemistry textbook. Students were required to generate self-explanations to facilitate the integration of new information into existing knowledge while reading the modeling-based text (Chi, de Leeuw, Chiu, & LaVancher, 1994).

Therefore, our hypothesis was that exposure to the explicit modeling-based text would result in students developing greater modeling competencies. The research addresses the following three questions:

1. What does student achievement related to the ideal gas law look like after students read the modeling-based text?
2. What do students' modeling competencies look like after they read the modeling-based text?
3. In terms of the modeling sequence, how do students' mental models of the ideal gas law change after reading the modeling-based text?

In this study, we defined modeling competencies as the thinking skills that students must exhibit to be successfully engaged in the modeling stages outlined above. More specifically, modeling competencies reveal the degree to which students choose appropriate model components (MS), join model components to generate a complete mental model (MC), test the internal or external consistency of their model (MV), use their validated model to analyze a target problem (MA), apply their validated model for solving problems in similar scenarios (MD), and revise their unvalidated model to form a modified model (MR). In this study, three categories of models were used: scientific models, mental models, and expressed models. Each term is defined as follows: scientific models refer to correct explanations of models used by chemists; mental models are students' explanations of specific phenomena; and expressed models include conceptual models, mathematical models, and submicroscopic models that learners used in this study. When a model consists of multiple components and the relationships between them, then these components might also form submodels for a specific phenomenon or law. For example, Boyle's law is a submodel of the ideal gas law.

METHOD

Participants

This study was conducted during the spring semester of 2010, and participants included fifteen 10th-grade students (ages 15–16 years) from a high school in New Taipei City, Taiwan. The participants all came from middle-class families. None of the participants had any previous experience in modeling-related learning or instructional activities. Participants were selected from two different classes taught by the same chemistry teacher by purposive sampling. The students were then classified into three groups: low, moderate, and high academic achievement in chemistry based on their chemistry grades during the fall semester of 10th grade in 2009. These three groups consisted of 8, 10, and 12 students, respectively. Within each of the three chemistry achievement groups, half of the students were randomly assigned to the treatment group. This resulted in four, five, and six students from the low, moderate, and high chemistry achievement groups, respectively, being assigned to the modeling-based text group. The students were not told they were in the experimental group, and they were requested not to discuss their participation or their text with each other outside of the study. During the research period, although the students continued their

fundamental chemistry course, the concepts presented in the modeling-based text (i.e., the ideal gas law) were not part of the students' regular chemistry curriculum. Therefore, exposure to the regular chemistry curriculum was not expected to influence the results of this study.

Materials

Reading Texts. The modeling-based text used in this study was a novel educational scientific text compiled by the authors for the purposes of this investigation. The text was based on *Eleventh Grade Chemistry*, a locally published high school textbook, which included a chapter on the ideal gas law. The modeling-based text covered Boyle's law, Charles and Gay-Lussac's law, Avogadro's law, the ideal gas equation, and descriptions of the modeling sequence and stages as set forth by Liu and Chiu (2010) and shown in Table 1.

The six modeling stages are described in relation to the specific gas law covered in each section of the modeling-based text (see Appendix A). In the beginning of each section of the modeling-based text, definitions and examples of scientific modeling practices are introduced and repeated in specific stages. The italics in the Appendices indicate the extra content in the modeling-based text compared to the above-mentioned textbook. Except for the additional content on interactions of gas particles in macroscopic phenomena and submicroscopic representations, the contents of both texts (modeling-based text and referent textbook) were the same for content validity purposes as judged by two science educators (physics and chemistry majors). The content and self-explanation questions at the end of each modeling stage were finalized after a pilot study conducted with three 10th-grade students at the same school.

The modeling-based text differed from the regular textbook in that the modeling-based text contained (a) explicit modeling stages for students to learn the full modeling process of developing a scientific conceptual model, (b) an emphasis on the limitations and applied scenarios of conceptual models (e.g., Boyle's law), and (c) a discussion of submicroscopic particles to explain the behavior of gases in macroscopic phenomena to provide students with a bridge for linking macroscopic and submicroscopic views.

Instruments

Gas Particle Concepts Assessment. There were 43 multiple-choice items in the assessment of students' main concepts, including particle view, rigid particles, particle motion is dynamic instead of static, random motion of gas particles, gas volume, and gas pressure involving conceptual understanding not algorithmic calculation (see Appendix B). The propositional descriptions are shown in Table 2. The students were able to complete the assessment within 50 minutes. The instrument was originally developed by Chiu (2012). The assessment instrument was reviewed and revised by two chemistry education professors, and revisions were made according to the comments from the reviewers. The assessment's Cronbach α was 0.92.

Semistructured Interviews of Modeling Competencies. We interviewed students to further assess their modeling competencies in relation to the six modeling stages (MS, MC, MV, MA, MD, and MR). The face validity of the questionnaire was verified by a team of three science educators who had at least 3 years of experience in the field of scientific modeling. A draft of the questionnaire was piloted at the same school with three

TABLE 1
Elements of Modeling Processes and Their Definitions of Each Step

Elements of Modeling Processes	Definitions[a]	Explanation of Each Step in the Modeling-Based Text
Model selection	To choose the appropriate objects (components)	There is a certain thought process when scientists create a scientific model. When they observe numerous similar events in life, they will begin to choose some abstract concept to interpret these events, a process known as model selection
Model construction	To build the related relations and structures of the chosen objects/components or basic models	The process of creating a scientific model based on the selected abstractions and including text descriptions, chemical formulas, and mathematical equations, is known as model construction
Model validation	To use the established relations and structures of a model to validate it to judge, test, or compare the internal or external consistency of the model	Judging if the text description, chemical formulae, and mathematical equations are reasonable, determining whether factors that will affect theory can be ignored, judging the feasibility of experimental methods, or considering alternative solutions is called model validation
Model analysis	To use the validated model for analyzing the problem and explaining its appropriateness (data calculation or reasoning)	After an established model is validated, it is necessary to analyze the data to identify the relationship between the variables and evaluate the differences between predicted and experimental results. This is known as model analysis and evaluation
Model deployment	Apply the validated model to solving problems in new contexts	The strengthening of the explanatory power of the model when the model is applied to other similar situations is called model deployment
Model recon-struction	To be aware of the failure of validated model and add or delete some objects/components to form a modification or transform it into a new model	Thinking of its limitations and re-adjusting the model when the model cannot be applied to either the required explanation or to other situations is called model reconstruction

[a]The definitions of the modeling stage are adapted from Liu and Chiu (2010).

TABLE 2
Main Concepts and Propositional Descriptions in the Gas Particle Concepts Assessment

Main Concept	Propositional Description	Number of Items
Particle view	Submicroscopic phenomenon: Gas exists in particle form The shortest distance between gas particles is in vacuum state	2
Gas particles are considered as rigid particles	The particle size is unaffected by external factors	3
Particle motions are dynamic instead of static	No energy will be lost in flexible collisions of gas particles, and all gas particles move continuously	4
The same quantity of gas particles moves in the same direction	Gas particles move in random directions in a sealed container Gas particles distribute at random in a sealed container	16
Definitions of gas volume and factors affecting gas volume	The space where gas particles move is the volume of gas Volume and mole count are in a positive relationship at the same temperature and same pressure Volume and absolute temperature are in a positive relationship at the same pressure and particle count	5
Formation of gas pressure and factors affecting pressure strength	Pressure generates as gas particles collide with the container walls Pressure and mole count are in a positive relationship at the same temperature and same volume Pressure and container volume are in an inverse relationship at the same temperature and same particle count Pressure and absolute temperature are in a positive relationship at the same volume and particle count The balloon temperature rises at fixed pressure; and pressure remains unchanged at equilibrium	13
	Total items	43

students and then revised by the above-mentioned three science educators. Two parallel interview protocols were finalized, and each contained two to six items for each stage. One questionnaire was administered as the pretest and midtest, and the second questionnaire was used as the posttest and delayed posttest (see Appendix C). Each interview lasted 50–60 minutes. All interviews were audio-recorded, and the audio files were transcribed verbatim into protocols for the analysis of students' modeling competencies.

Research Procedure

Prior to the experimental phases, all 15 students participated in two 20-minute training sessions over a 1-week period to learn how to read the required text and use the recorder to record their answers to the self-explanation questions. The themes in the modeling-based text in order were Boyle's law, Charles and Gay-Lussac's law, Avogadro's law, and the ideal gas equation. These topics were covered in 10 weeks. During the first week, a pretest, including the modeling competencies interview and the gas particle concepts assessment, was conducted with the participants prior to reading the modeling-based text. Then, once a week, participants were asked to read their texts while under the supervision of a research team member. Before each reading activity, each supervisor described the instructions of the reading task. Supervisors recorded the time each student spent on the reading task and also recorded any comments or queries by the students. The supervisors were present only to ensure the students were engaged in the reading task, and they did not interfere with the students' reading or associated activities (e.g., students' self-explanations). A midtest, including an interview and a formative assessment, was arranged during Week 4 . A posttest, including an interview and multiple-choice test, was conducted after the reading of all texts during Week 7. Then, a delayed posttest was administered 3 weeks later.

When reading the texts, the students were requested to follow the instructions for the questions in the text to generate self-explanations to facilitate text understanding (Chi et al., 1994). Texts were divided according to the modeling stages adapted from Liu and Chiu (2010). The self-explanation questions were immersed within the descriptions of the modeling stages in the modeling-based text and were meant to help students check their understanding as they read about each stage in the modeling sequence. The students reading the modeling-based text read each passage once and spent 45 minutes on average (range of 25 minutes to 1.5 hours) reading their text. At the start of each session, students were requested to take 5 minutes to review the previous reading before continuing with the next.

Data Collection

A mixed-method was adapted in this study. Quantitative data were collected from the results of the gas particle concepts multiple-choice items on the pretest, posttest, and delayed posttest. Qualitative data were collected from the semistructured interviews of modeling competencies in the pretest, midtest, posttest, and delayed posttest.

Coding for Modeling Competencies. The rubric scheme used in this study, modeling competence analytic index (MCAI) was established by Chang and Chiu (2009) and modified from the standards of the Structure of Observed Learning Outcome (SOLO) taxonomy proposed by Biggs and Collis (1982), falling into L_0–L_5. L_0 refers to unanswered items or the answer, "I don't know." L_5 refers to the correct scientific explanation for the ideal gas law (see Table 4). The SOLO taxonomy has been used successfully to classify the progression of students' responses (Branbrand & Dahl, 2009; Minogue & Jones, 2009). The higher the level of student performance, the more students demonstrated modeling competencies and made correct connections between science concepts.

There were three steps involved in the coding of students' interview responses (see Tables 3 and 4). Step 1 was to confirm the characteristics of the modeling competencies that students had to exhibit to be successfully engaged in the modeling stages (MS, MC, MV, MA, MD, and MR). For example, in analyzing the behavior of gas bubbles underwater, when students considered the possible variables, such as pressure and volume, we coded students as choosing appropriate components to explain the bubbles' change in size after a

TABLE 3
Description and Examples of the Response for Each Level

Responses of Each Level	Description of Response Main Level	Coding Example
L_0 Prestructural	Unanswered or "I don't know"	Not sure or I don't know
L_1 Unistructural	Description of a single related factor	Answers with *P*, *V*, *n*, or *T*
L_2 Multistructural	Description of the qualitative relationship between two related factors	Answers with two related factors, such as the qualitative description of *P*, *V*, or *P* is larger, when *V* is smaller
L_3 Relational	Description of the mathematical relationship between two related factors	Answers with $PV = K$, with description that pressure is the normal force of unit area, and volume is the space occupied by the balloon
L_4 Extended abstract	Description of the mathematical relationship between three related factors	Answers with $PV = KT$, with description that pressure is the normal force of unit area, volume is the space occupied by the balloon, and temperature is the air temperature measured with a thermometer
L_5 Scientific explanation	Full description of the ideal gas equation	Correct scientific explanation

Abbreviations: *P*, pressure; *V*, volume of gas; n, numbers of particle; *T*, absolute temperature.

diver exhaled and ascending bubbles formed (coded MS). If students were able to use the appropriate variables, such as $PV = K$, to describe the phenomenon, we coded them MC. Student responses that involved designing a thought experiment to test the consistency of relationships among variables were coded MV. Responses that used the validated mental model, such as $PV = K$, to analyze the problem were coded MA, and responses that applied the validated model, such as $PV = K$, for solving problems in similar scenarios were coded MD. When students encountered an additional variable in the interview questions, students revised the original relationships between the variables to form a modified model, such as from $PV = K$ to $PV = KT$ or $PV = nRT$ (coded MR).

Step 2 was to confirm the main level in terms of the grades in the SOLO taxonomy in each modeling competency from Step 1. The main level was classified by the relationship among macroscopic factors in each modeling competency, with the exception of MS which was coded by the number of variables in students' responses. For example, one student used pressure, temperature, and volume to describe a balloon flying toward the sky and used $PV = K$ or $V = KT$, which we coded as L_3. Once the relationships between variables increased, the main level also increased. That is, a main level increased after an explanation increased, such as from grade L_2 to L_3, L_4, or L_5. For instance, the response of $PV = KT$ was encoded L_4.

Step 3 was to confirm the sublevel in each main level from step 2. The sublevel was classified by the interactions between the submicroscopic representations and macroscopic phenomena. For example, a student who only used $PV = K$ to describe a balloon flying toward the sky, who could not define pressure and volume with submicroscopic understanding, and who could not use behavior of gases to explain why was coded $L_{3,1}$. If the student

TABLE 4
Description and Examples of Response Subhierarchy With L_3 Response Main Hierarchy

Responses of Sublevel	Description of Responses of Sublevel	Coding Example
1	Macroscopic descriptions	Answers with $PV = K$, with description that pressure is the normal force of unit area, and volume is the space occupied by the balloon
2	Including the macroscopic and submicroscopic descriptions, except explaining macroscopic phenomena with submicroscopic phenomena	Answers with $PV = K$, with description that pressure is the normal force of unit area; volume is the space occupied by the balloon and the space where air particles move
3	Including the macroscopic and submicroscopic descriptions, with explanation of macroscopic phenomena with submicroscopic phenomena	Answers with $PV = K$, with description that pressure is the normal force of unit area and generated by the collisions of air particles against the container walls; when external pressure reduces, the force of collisions of air particles against the container walls inside the balloon changes as a result; comparatively, the distances for air particles increase, the space for air particle motion increases, and the balloon volume increases

Abbreviations: *P*, pressure; *V*, volume of gas.

described the definitions of pressure and volume with submicroscopic understanding and $PV = K$, we coded the response $L_{3.2}$. A student who described not only the definitions of pressure and volume but $PV = K$ with submicroscopic understanding to explain why a balloon flies toward the sky, was coded as $L_{3.3}$. When coding students' responses to the questionnaire, we discovered that high school students could describe the relationships between both factors in a substantial way because they had learned the concepts of Boyle's law in junior high school but not the ideal gas law. Therefore, considering other studies with codes similar to SOLO (Minogue & Jones, 2009) and that the sequence of mathematical model construction is from qualitative description to quantitative analysis (Lesh & Doerr, 2003), students who could answer two relevant factors or their qualitative relationships were coded as L_2 in this study. The interrater agreement of the modeling competencies was conducted on 20% of the respondents (three students from the modeling group) by two coders. The agreement was 85%, and disagreements were resolved through discussion.

Data Analysis

This study adapted a mixed method to analyze the data collected. The quantitative data included assessment of the students' gas particle concepts, the transformation of students' protocols of modeling competencies into scores, and the relationships of all students'

conceptual and modeling competency scores. The qualitative part included the students' protocols from the interviews on modeling competencies. The gas particle assessment consisted of multiple-choice items, and results were expressed as percentages. The results of the modeling competency interviews were transcribed into protocols and then coded from 0–5 according to the rubric (see Table 5). We evaluated modeling competencies by scoring 0–4 to represent the level of competency in making relationships between the components that comprise the ideal gas law. Except for model selection, students' modeling competencies were based on the number of variables and the number of relationships between variables used for solving problems. They were first scored with the main level, from L_0–L_5, where L_0 was 0 points, L_1 was 0–1 points, L_2 was 1–2 points, etc. The maximum score for this level was the number of the main level. For example, the maximum score of L_2 was 2 points. Then, by converting the score of the sublevel, the score distribution for that interval was formed. Take code 3.1. for example, the score was $2 + 0/3 = 2.00$ points; the score in code 4.2 was $3 + 1/3 = 3.33$ points. The coding scheme of model selection started with students selecting the components of the model and then explaining the variables (P, V, n, and T) of the ideal gas law. If a student explained the phenomenon with only one variable, that student scored 0.25, 0.5 for two variables, 0.75 for three variables, and 1.00 for four variables. This way, the answers for the individual modeling stages within each modeling competency item were translated into a quantifiable value for statistical purposes. SPSS 17.0 was used for statistical analysis. As there were 15 participants, nonparametric statistical techniques (Wilcoxon test) were applied to the quantitative data.

RESULTS

In this section, both the quantitative and qualitative findings are described to show how students performed in terms of their conceptual understanding and modeling competencies. In particular, we present the results for student 03 as a case example of how the students went through each step of the modeling sequence and formulated mental models of the ideal gas law. Table 6 provides the data and analysis for each research question.

Finding 1: The Impact of the Modeling-Based Text on Students' Understanding of the Ideal Gas Law

As shown in Table 7, the Wilcoxon test was applied to analyze the number of correct responses on the pretest (35%, $SD = 9.38$), posttest (66%, $SD = 19.93$), and delayed posttest (68%, $SD = 12.39$) related to gas particle concepts for a one-group comparison. Results showed that the differences from pretest to posttest ($z = -3.35$, $p = .001 < .05$) and from pretest to delayed posttest ($z = -3.41$, $p = .001 < .05$) were significant. The students who read the modeling-based text gave more correct responses to the test items assessing gas particle concepts and demonstrated greater knowledge *after* reading the modeling-based text compared to before reading the text. This effect was still present 3 weeks after the intervention ended.

Finding 2: The Impact of the Modeling-Based Text on Students' Modeling Competencies During Interviews

The students' protocols from the interviews were converted into scores according to the coding scheme of modeling competencies. Modeling competencies for each student were measured in terms of MS, MC, MV, MA, MD, and MR. The Wilcoxon test was applied to test the overall modeling competencies of the pretest (6.84, $SD = 1.62$), midtest

TABLE 5
Coding Examples and Scoring

Level	Code	Macroscopic	Symbolic	Submicroscopic	Score
0	0		No answer/"I don't know"		0
1	Coding example	Answers with P	Pressure is the normal force of unit area	Pressure generated by the collisions of air particles against the container walls	
	1.1	O	X	X	0.25
	1.2	O	O	X	0.50
	1.3	O	O	O	0.75
2	Coding example	Answers with P and V or qualitative description: P is larger when V is smaller	Pressure is the normal force of unit area; volume is the space for air particle motion; and their qualitative relations	Pressure generated by the collisions of air particles against the container walls	
	2.1	O	X	X	1.00
	2.2	O	O	X	1.33
	2.3	O	O	O	1.67
3	Coding example	Answers with $PV = K$, with description: Pressure is the normal force of unit area, and volume is the space occupied by the balloon	Pressure generated by the collisions of air particles against the container walls; volume is the space for air particle motion	When external pressure increases, the force of collisions of air particles against the container walls inside the balloon increases as a result; comparatively, the distances for air particles reduces, the space for air particle motion reduces, and the balloon volume reduces	
	3.1	O	X	X	2.00
	3.2	O	O	X	2.33
	3.3	O	O	O	2.67
4	Coding example	Answers with $PV = KT$, with description that pressure is the normal force of unit area or temperature is the air temperature measured with a thermometer	Pressure is generated by the collisions of air particles against the container walls, volume is the space for air particle motion, and temperature is the average of kinetic energy for air particles	The average kinetic energy of gas reduces as the temperature drops; velocity of particle motion reduces, and collisions against the container walls reduce; balloon internal pressure reduces; comparatively, the distances for air particles decrease, particle motion space reduces and balloon volume reduces	

TABLE 5
(Continued)

Level	Code	Macroscopic	Symbolic	Submicroscopic	Score
	4.1	O	X	X	3.00
	4.2	O	O	X	3.33
	4.3	O	O	O	3.67
5	5	Correct scientific explanation	Correct scientific explanation	Correct scientific explanation	4.00

Abbreviations: *P*, pressure of gas; *V*, volume of gas; *n*, number of particles; *R*, ideal gas constant; *T*, absolute temperature; O, students' responses were consistent with the codes; X, students' responses were not consistent with the codes.

TABLE 6
Data Set and Analysis for Each Research Question

	Content	Tool	Analysis
Question 1	What is student achievement related to the ideal gas law changed after reading the modeling-based text?	•Gas particle concepts assessment	•Wilcoxon test
Question 2	What is student modeling competencies changed after reading the modeling-based text?	•Semistructured interviews •The rubric scheme, modeling competence analytic index	•Wilcoxon test
Question 3	How do the students construct their mental models of the ideal gas law in the modeling process?	•Interview data	•Verbal report

TABLE 7
Pretest, Posttest, and Delayed Posttest of Science Concepts

Test	Mean of Correct Responses (%)	Standard Deviation
Pretest	34.88	9.38
Posttest	65.89	18.93
Delayed posttest	68.22	12.39

(10.18, $SD = 2.67$), posttest (13.16, SD = 3.29), and delayed posttest (14.92, SD = 2.88) for one-group comparison (see Table 8). Results show that the differences in the correct responses between pretest and midtest ($z = -3.24$, $p = .001 < .05$), pretest and posttest ($z = -3.32$, $p = .001 < .05$), and pretest and delayed posttest ($z = -3.35$, $p = .001 < .05$) were significant. The results indicate that the modeling group had better performance on the midtest, posttest, and delayed posttest than pretest.

The scoring factor ($X - 0.33$) of interaction for the macroscopic phenomena and submicroscopic representations was eliminated to prevent attribution interference as a result of explaining the differences for the modeling-based text. The Wilcoxon test was applied to retest the modeling competencies with significant differences as shown in Table 9. The

TABLE 8
Overall Modeling Competencies

Test	Mean	Standard Deviation
Pretest	6.84	1.62
Midtest	10.18	2.27
Posttest	13.16	3.29
Delayed Posttest	14.92	2.88

results show that after eliminating the potential factors in interaction with macroscopic phenomena and submicroscopic representations, the differences in MS, MC, MV and MR between pretest and midtest and MS, MC, MV, MA, MD, and MR between pretest and posttest and pretest and delayed posttest were still significant.

Finding 3: How Students Elaborated Their Mental Models of the Ideal Gas Law Over Time

We examined how students' mental models explained the bubbles moving around and how they tested their models during the readings and interviews.

Pretest. As shown in the results from the modeling competency analysis, prior concepts about the ideal gas law for students were based on the qualitative inference of two variables. For example, students at pretest believed the pressure outside the balloon increased and then the volume inside the balloon decreased. When students were unable to recognize the internal limitations of the model in MV, they usually considered the control variables of the ideal gas equation as limitations or did not know what the model limitations were. Large protocols were coded in each modeling stage to examine students' modeling competencies and mental models. Here, our analysis stressed students' MC and MV competencies. Through analyzing students' discourse of MC and MV, we were able to confirm students' mental models and the differences in their key modeling stages. While not representative of all the students in the group, the case described below does clearly demonstrate student' mental models and how the case tested the reasonableness of mental models during reading activities and interviews. The following conversations with S03 (a student who read the modeling-based text) illustrate the student's intent to explain the changes in volume through the concepts of pressure and temperature and to establish and validate the mathematical relationship. When asked about the factors affecting the moving bubbles, S03 answered that "pressure" or "temperature" affected the "volume" of bubbles. For example, "Pressure will affect the volume of the bubbles" (S-03-pretest). The students were asked to design a thought experiment to test the reasonableness of their mental models and to specify the limitations of their experiment. Student S03 said:

S03: Bubbles in the aquarium.
S03: It is a matter of water pressure, and it needs to calculate well, just at the height that bubbles break, otherwise, bubbles will attach to the glass without breaking. If we need to take the bubbles, we may need a pipette, because there is air inside, this may bring errors to the experiment. (modeling-based group-S-03-pretest)

Before the reading activity, the student, S03, had no idea about Boyle's law. In her self-explanations at the first reading activity, S03 said, "I never heard about Boyle's law before this reading activity." This student was only able to develop mental models of the ideal gas law through qualitative means rather than the quantitative equation. Although the

TABLE 9
Modeling Competencies of Both Groups After Eliminating Potential Interactions Between Macroscopic Phenomena and Submicroscopic Representations

Comparison	Modeling Stage	Rank	Number	Total of Rank	z	p
Pretest and midtest	MS	–	2	5	–2.68	.007**
		+	10	73		
	MC	–	1	4	–2.59	.010**
		+	10	62		
	MV	–	2	14	–1.97	.049*
		+	10	64		
	MA	–	9	50	–0.58	.562
		+	6	70		
	MD	–	2	9	–1.89	.058
		+	8	46		
	MR	–	1	4	–2.59	.010**
		+	10	62		
Pretest and posttest	MS	–	0	0	–3.24	.001***
		+	13	91		
	MC	–	1	2	–3.19	.001***
		+	13	103		
	MV	–	1	2	–3.19	.001***
		+	13	103		
	MA	–	5	20	–2.29	.022*
		+	10	100		
	MD	–	1	3	–2.84	.005**
		+	11	75		
	MR	–	1	2	–3.19	.001***
		+	13	103		
Pretest and delayed posttest	MS	–	0	0	–3.34	.001***
		+	14	105		
	MC	–	1	1	–3.37	.001***
		+	14	119		
	MV	–	1	1	–3.41	.001***
		+	14	119		
	MA	–	3	6	–3.10	.002**
		+	12	114		
	MD	–	1	3	–3.26	.001***
		+	14	117		
	MR	–	1	1	–3.37	.001***
		+	14	119		

*$p < .05$; **$p < .01$; ***$p < .001$.
Abbreviations: +, positive ranks; –, negative ranks; MS, model selection; MC, model construction; MV, model validity; MA, model analysis; MD, model deployment; MR, model reconstruction.

student was able to mentally design an experiment to test her mental model, the student was unable to explain the restrictions of the ideal gas equation under certain scenarios, except for insignificant environmental elements.

The First Reading Activity. In her self-explanations following the first reading activity about Boyle's law at the stage of MC, S03 said, "There are two variables, pressure and volume, in Boyle's law; the pressure increases, then the volume decreases." The presentations of MR from the modeling-based group are represented by this response, "We usually misunderstand Boyle's law, which $P \times V =$ constant. The law works at fixed temperature and number of molecules." At the end of the reading, S03 modified her explanations of MR, "Boyle's law demonstrates that inverse relationship between pressure and volume under the controlled particle numbers and temperature conditions." The student demonstrated an understanding of the relationship between the four variables and that Boyle's law was the product of pressure and volume when the remaining two variables are held constant.

The Second Reading Activity. In their self-explanations following the second reading activity about Charles and Gay-Lussac's law at the stage of MC, S03 said, "At the conditions of controlled pressure and numbers of molecules, the temperature increases; then the volume decreases." At the end of the reading, S03 modified her explanations and demonstrated her changed thinking in her self-explanation of MR. She said, "I did not know Charles and Gay-Lussac's law until I read the text," and "the unit of temperature in Charles and Gay-Lussac's law must be Kelvin temperature scale." Similarly, at the stage of MC, S03 said, "At the conditions of controlled pressure and number of molecules, the volume should be proportional to the temperature," and "the temperature scale is Kelvin rather than Celsius scale." After reading the modeling-based text, the student could describe the relationship among the four variables and how Charles and Gay-Lussac's law could be influenced by volume and absolute temperature at fixed pressure and particle number.

Midtest. After the two previous reading activities, on the midtest students who read the modeling-based text explained the mechanism of macroscopic phenomena with submicroscopic gas particle behavior. Some of them still explained the mechanism of submicroscopic particle motions with macroscopic phenomena. When asked on the midtest about the factors affecting the moving bubbles, S03 wrote down the mathematical relationship $PV = K$, and explained, "If it moves upwards, water pressure gradually reduces, the air pressure inside differs from the pressure outside. In order to maintain the same [pressure], it must increase its volume, because it cannot change the particle count, therefore it increases its volume," and "When it's hot, the particles move faster." S03 not only designed a thought experiment, bubbles in an aquarium, to test the reasonableness of her mental model, but she specified certain criteria and restrictions, "the same kind of water," in the design of the experiment. In fact, S03 designed a thought experiment, but she stressed a noncritical factor.

By the midtest, students had finished the reading activities related to Boyle's law and the Charles and Gay-Lussac's law. The modeling group student, S03, was able to establish the mathematical relationship between pressure and volume but unable to consolidate the relationship that formed the three variants. Meanwhile, S03 was also able to propose the limiting factors related to the ideal gas equation to design an experiment to verify and validate the model she constructed.

The Third Reading Activity. In her self-explanations following the third reading activity about Avogadro's law at the stage of MC, S03 said, "There are four variables, pressure, volume, numbers of molecules and temperature in Avogadro's law. The number of molecules increases, then the volume increases, too." At the end of the reading, S03 demonstrated her

explanations of MR to elaborate the Avogadro's law. "The volume of a balloon comes from the number of molecules rather than molecular weight of gas. For example, the volume of hydrogen will be larger if two balloons are packed with four moles of H_2 and two moles of Cl_2 gas, respectively."

The Fourth Reading Activity. In her self-explanations to the fourth reading activity about the ideal gas equation of MR, S03 said, "I would consider that the volume should be proportional to the number of molecules rather than the density of gas be proportional to molecular weight, if the conditions are at the controlled temperature and pressure. After reading, I know that it is not a contradiction to have $PV = nRT$ modified to $PM = dRT$" and "understanding under what conditions the behavior of real gases will be close to the ideal gas."

Posttest. After students finished all the reading activities, the students who read the modeling-based text were able to carry out critical thinking on the tests designed to verify the established model and describe the restrictions associated with using the model. When designing experiments for testing their models, the students could gradually consider viable experiments independently, including considering the limitations supported by the mathematical model. Student S03 was able to establish the mathematical relationships among pressure, temperature, and volume during the posttest ($PV = KT$) and design an experiment to verify the model related to the ideal gas equation she constructed.

Researcher:	Please judge, when both temperature and pressure reduce, will volume expand or reduce?
S03:	Expand. Regardless of pressure changes, the higher the altitude, the smaller the pressure, and the lower the temperature. However, regardless of the changes, the quotient must be larger than the original value. Because the difference in this (pressure) reduction is bigger than that (temperature) reduction, the upper one is bigger (then student wrote down the formula, $V = kT/P$).
Researcher:	Please specify the experiment limitations.
S03:	Limitation, the balloon cannot be blown.
Researcher:	What else?
S03:	Then, particles inside the balloon cannot be changed, gas particles, molecule count! (modeling-based group-S-03-posttest)

Delayed Posttest. On the delayed posttest conducted 3 weeks after the final reading session, the students who read the modeling-based text progressed from establishing the relationship between two variants to three variants. These students used their submicroscopic understanding of gas particle behavior to explain the macroscopic phenomenon of the theoretical mechanism, as well as make reasonable deductions on the limited criteria in the experiment through the factors associated with the ideal gas equation. When the students were asked to design experiments for validating their models, the students were able to adapt their tentative models to incorporate related factors (e.g., pressure, volume, temperature, number of particles) with the ideal gas equation to justify the model's reasonableness.

The students who read the modeling-based text were able to establish accurate mathematical relationships among the three variants of pressure, temperature, and volume ($V = kT/P$). "Pressure and volume are in ... an inverse relationship, temperature and volume are in a positive relationship. Pressure may change the volume; therefore, the smaller the

pressure, the larger the volume." These students interpreted macroscopic phenomena with representations of submicroscopic symbols and designed experiments using the internal factors of the ideal gas equation model as the restrictions: "Limitation, the balloon cannot be blown," and "Particles inside the balloon cannot be changed, it means numbers of gas particles are the same."

Researcher:	Please specify what is the meaning of $V = kT/P$?
S03:	I think temperature drops slower and pressure drops faster. Overall, the value will increase. Therefore, the volume increases.
S03:	Because when pressure drops, temperature will also drop. Maybe pressure drops faster than temperature. When air pressure drops from 1 atm to 0.5 atm, and temperature may drop by 5 degree. After division, this is still bigger. Therefore, its value is bigger in general, and the volume increases.
Researcher:	Please explain your description in Item 5 from the viewpoint of gas particles.
S03:	When the balloon rises, the air pressure reduces, and the balloon pressure increases. To balance the pressure, gas particles will hit against the container walls, and volume increases (external). Temperature then ... also drops, but (when compared with pressure) less .... Accordingly, the gas motion inside the balloon should be slower, but when the air pressure outside reduces, the pressure inside the balloon increases. This way, gas particles will continue to hit the container walls. As the hits increase, volume increases. (modeling-based group-S-03-delayed test)

In this study, students were requested to design a mental experiment to validate the predictability of their established models. Instructing students on the relevant modeling knowledge resulted in more effective problem solving with different scenarios through transfer of models after students had evaluated the models' correctness and reasonableness. That is to say, by allowing students to evaluate the models they established and validate the models' reasonableness, students recognized the complete modeling processes of scientific concepts, recognized the limitations of established models, and learned more accurate scientific concepts.

DISCUSSION AND CONCLUSIONS

This study explored and evaluated the effectiveness of modeling-based text in facilitating students' conceptual understanding and modeling competencies about the ideal gas law. The results showed a significant improvement in conceptual learning and modeling competencies by the students who read the modeling-based text. We argue that the affordances of the modeling-based text, emphasizing explicit modeling processes, prompts students to utilize modeling competencies and science concepts and leads to increased systems-level conceptual understanding. As shown in Figure 2, the text covered the central scientific concepts of the ideal gas model and was designed to include the modeling stages (i.e., MS, MC, MV, MA, MD, and MR) to help students to learn the ideal gas law by means of reconstructing scientific models. During the process, students were gradually introduced to the meaning and applications of the law, which provided opportunities to reconstruct their mental models if those models were invalidated. We argue that teaching students the scientific modeling stages facilitated students' learning of the ideal gas law. In what follows, we discuss and conclude our findings regarding the design of an innovative modeling-based text and highlight the value of this work.

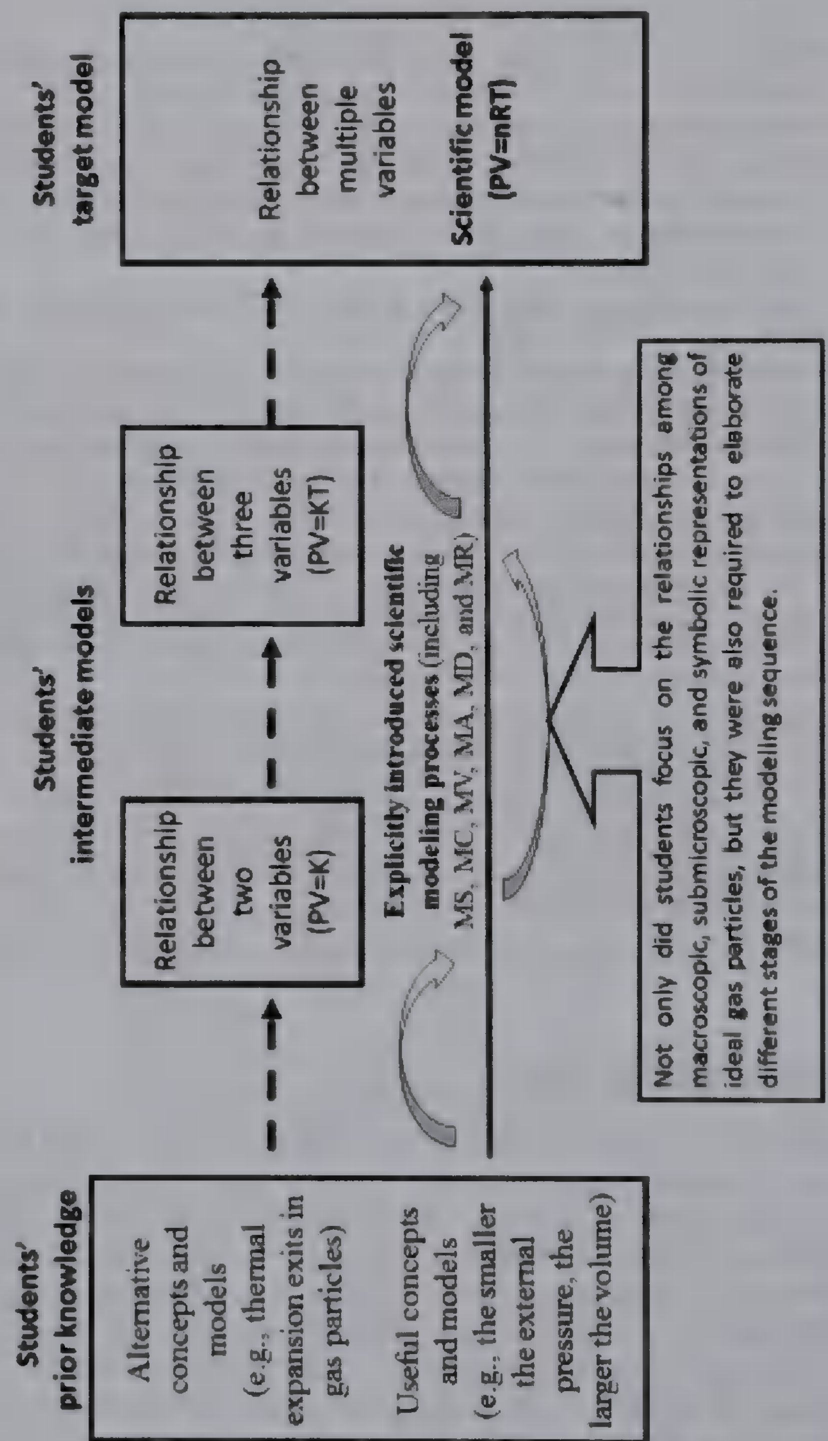


Figure 2. The development of students' mental models and modeling competencies.

Modeling-Based Text Promotes Modeling Competencies and Concept Learning

A model is a limited version of its target. Scientific modeling is a means of revising and reconstructing scientific models. School science should provide not only the scope, functions, and limitations of conceptual models but should also engage students in understanding how a conceptual model is constructed and developed to facilitate their understanding of scientific phenomena. On the basis of this premise, we designed a modeling-based text that stresses the modeling sequence and stages related to the ideal gas law. This study leads to better understanding about the impact on science learning of a novel version of science text and the subsequent changes in students' conceptual knowledge and modeling competencies. Students were able to engage in the modeling process through the reading activities including describing variables, generating a relationship between variables, describing an experiment to examine the reasonableness of their constructed models, and applying the model to similar contexts. Six modeling stages (MS, MC, MV, MA, MD, and MR) were introduced in the modeling-based text to help students to recognize the limitations of their models. Through reading the three submodels of the ideal gas law in the modeling-based text, the students realized the limitations of Boyle's law, Charles and Gay-Lussac's law, and Avogadro's law, and reorganized factors (pressure, volume, number of particles, and temperature) to form the ideal gas equation. In addition, the students distinguished properties of ideal gas from real gas to modify the ideal gas law.

As demonstrated in the literature, modeling-based teaching can help students to acquire relevant knowledge (Dori & Kaberman, 2012; Halloun, 1996), but as this study demonstrated, students can also acquire relevant knowledge through reading activities. A modeling-based text provides an opportunity to promote students' modeling competencies and conceptual learning via reading activities. The modeling-based text described explicit modeling stages to help students learn the full sequence of scientific modeling. The students in this study performed better after reading the materials we provided. The modeling-based text not only facilitated students' understanding of how to calculate but also helped them understand how to apply mathematical calculations and overcome limitations in different scenarios to apply the same model to a different context where one variable was changed. Our results differ from some of the research on refutational or conceptual change text. Refutational or conceptual change text is meant to overcome conceptual limitations, whereas the modeling-based text acknowledges of the limitations of the scientific model. The modeling-based text extends the network of connections between concepts using system-wide thinking to generate, modify, revise, and reconstruct the models students constructed rather than only revising a specific misconception.

Develop Competency in Recognizing Multivariable Relationships and Transferring This Knowledge to the Ideal Gas Law

Previous research on the ideal gas law has suggested that in-depth conceptual understanding, involving connections between chemical principles and the various chemical representations that describe phenomena, increases students' concept learning (de Berg, 1989; Niaz, 1995; Niaz & Robinson, 1992). The results from our study demonstrate how students can improve their conceptual understanding and improve their modeling competencies. Furthermore, the modeling-based text engaged students in the reading activity by allowing them to validate and see the interconnections of the concepts associated with the ideal gas law. The way that students reasoned about the use of multiple models was similar to Gericke et al. (2013) even though they did not use the term model. When students did not recognize and incorporate the components of scientific models, as shown in the

interviews, students had difficulty combining different conceptual models. For example, students did not understand p, v, n, T, $PV = K$, $V = KT$, or $n = KT$. They did not combine $PV = K$, $V = KT$, or $n = KT$ into $PV = nRT$. This study differs from the investigation by Gericke et al. that explored students' ability to discern conceptual variation through reading textbooks. Our study focused on not only the recognition of conceptual variation but also on the relationship between variables in the students' mental models. Furthermore, this study designed a novel modeling-based text to help students integrate multiple models and understand the relationships between them. The students' performances of modeling competencies were identified from the relationships between the variables.

Instead of oral instruction, this study adopted a self-explanation strategy for 10th graders. Students' explanations, showing the tendency to return to their old and trusted concepts, were in some cases better after the reading than in the interviews because students could check their answers according to the text during the reading activities. The evolutionary path of the ideal gas equation model began from a qualitative description. It further developed toward the relationship between two variables and gradually ended with multivariable relationships. That is to say, initially students held qualitative descriptions of variables, then they constructed mental models from the model $PV = K$ ($V = KT$ or $V = nK$) to the model $PV = KT$, eventually forming the scientific model $PV = nRT$. Students who used the modeling-based text developed multivariable relationships and constructed correct mental models.

The SOLO Rubric to Confirm Students' Responses About the Relationships Between Variables

Some studies define competency as the level of sophistication in students' scientific concepts (Dori & Kaberman, 2012; Schwarz et al., 2009), and some stress the features of how to generate, revise, and evaluate with internal or external models (Gilbert, 2005; Wang & Barrow, 2011). We argue that competency in generating, revising, and evaluating with internal or external models *cannot separately* be measured by characterizing features or levels of sophistication of scientific concepts but must instead integrate and stress both how students form patterns of behavior and how they utilize specific concepts in specific contexts. In the present study, we not only characterize modeling competencies as students' performances during different stages of the modeling process but also emphasize how to use modeling competencies to generate, modify, and reconstruct a false model into a correct scientific model in a scientific text that explicitly demonstrates the scientific modeling process.

In contrast to Wang and Barrow (2011) and other studies on the rubrics of modeling competencies that describe the levels in terms of high, moderate, and low categories, the rubric in these studies *could not clearly* show the interaction between the components in the model. This rubric was adapted by Chang and Chiu (2009) and integrated the SOLO taxonomy and the corresponding rule of scientific explanation between a scientific theory and phenomena as a modeling competency analytic index, which showed the interaction of the compositions in a conceptual model and distinguishing the five levels of specific modeling competence. Unlike other SOLO studies, in this study we took students' prior knowledge into consideration and scored the qualitative relationship of two variables in the ideal gas law as L_2 rather than L_3 to avoid overestimation. By distinguishing between the levels of modeling competencies, we were not only able to promote students' competence in transferring their understanding of different models, we also expanded their competence in dealing with parameters of the ideal gas equation model.

Implications

Through integration of additional supports in the modeling-based text, students will be able to identify concepts in scientific models that were not previously known. This study demonstrates the effectiveness of modeling-based text. The major implication from this study is how to improve the structural sequences of science textbooks. This study suggests that there is a critical need to build support structures to help students understand the process of creating a scientific model, choosing an abstract concept to interpret a specific event, judging the feasibility of experimental methods, considering alternative solutions, analyzing the data to identify the relationship between the variables, applying to similar situations, and thinking about the limitations of models, as this process could lead to a better understanding of scientific models. In this study, the complete modeling sequence related to the ideal gas law was blended with the modeling-based text in an explicit teaching approach. The results show that explicit and complete modeling stages can help students to test the established model and recognize model limitations, thus allowing students to apply these models to different scenarios. Based on the lack of scientific modeling texts (Morgan & Morrison, 1999; Schwarz et al., 2009), the results of this study provide a reference for the preparation of relevant textbooks.

Generalizability, Validity, and Limitations

The findings in this study leave us with some possible limitations and raise questions about the generalizability and validity of the results due to the sampling and methodology used. First of all, we argue that modeling competencies are not only necessary for learning about the ideal gas law in chemistry but also for learning about science in general. Concepts from chemistry, biology, earth science, and physics consist of components and their relationships; therefore, modeling as conducted by scientists is relevant to all of these fields. A scientific phenomenon is usually too complex and abstract to be described completely. Scientists usually use different kinds of representations to represent what they think. For example, chemists build macroscopic, submicrosopic, and symbolic representations of chemical knowledge to communicate and share with others. We believe this study supports the use of modeling-based text not only for learning the ideal gas law but for learning science in general across disciplines.

Second, students were requested not to discuss their participation with each other outside of the reading activities, and students said they did not discuss their experience with other students. However, if students did discuss their participation this could be a threat to the validity of the study.

Third, students who participated in this study were assigned randomly into the experimental group, but rather were selected from specific classes. It is possible that our results reflect a certain limitation given the homogeneous nature of the student pool.

Testing models to recognize their limitations can help students to reaffirm relevant concepts through metacognition (Blank, 2000). The metacognitive competency of students affects the development of their modeling competencies. Further studies should focus on developing students' modeling competencies, in particular in validation and deployment, and provide more opportunities for students to develop their modeling competencies in school science classes.

The authors would like to thank the reviewers for their helpful comments and suggestions on this manuscript.

APPENDIX A: SELECTED MATERIAL OF THE MODELING-BASED TEXT

Modeling Stage	Modeling-Based Text
Model Selection	***[model selection]:*** *There is a certain thought process when scientists create a scientific model. When they observe numerous similar events in life, they will begin to choose some abstract concepts to interpret these events, a process known as "model selection."* [a]Based on the three gas laws from sections 2–4, we now know that we can describe gases with physical attributes of pressure, absolute temperature, volume, and the number of moles and construct a formula for ideal gases. *[Explain the problem: Please express your thoughts in detail to the voice recorder.]* ※*Q1 Please describe how "model description and selection" was conducted?* ※*Q2 What are the variables of the ideal gas law? What are the possible relationships among the variables?*
Model Construction	[***model construction]***: *The process of creating a scientific model using selected abstractions, text descriptions, chemical formulas, and mathematical equations is known as "model construction".* The three laws from the previous three sections can be integrated as follows: Boyle's law: $V \propto 1/p$ (constant temperature and constant volume) Charles's law: $V \propto T$ (constant pressure and constant volume) Avogadro's law: $V \propto n$ (constant pressure and constant temperature) Combining the three laws, we have formula 5-1 $v \propto \frac{nT}{P}$ (5-1) $v = R\left(\frac{nT}{P}\right)$ (5-2a) Formula 5-1 can be rewritten as equation (5-2 a) Alternatively written as $PV = nRT$.... (5-2b) 5-2b has been called the ideal gas equation, where R is the universal gas constant. The value of the gas constant R can be calculated when STP under the standard molar volume of the gas is taken into the ideal gas equation, $R = \frac{PV}{nT} = \frac{(1.0000\ \text{atm})\ (22.414\ \text{L})}{(1.0000\ \text{mol})\ (273.15\ \text{K})} = 0.082057 \frac{\text{L atm}}{\text{mol K}}$ the values of R often take two significant figures in general calculations, R = 0.082 L atm/mol K. The amount of the substance of the gas in moles equals the mass of the gas (W) divided by its molecular mass (M). The ideal gas equation can be written as formula 5-3 $PV = \frac{W}{M}RT$ (5-3) 5-3 can be transformed into 5-4 $PM = \frac{W}{V}RT$ (5-4) W/V is the density of air that can be written ≡ d. $PM = dRT$ (5-5) Formula 5-5 shows that, at constant temperature and constant pressure, the density of the gas is proportional to its molecular weight. *[Explain the problem: Please express your thoughts in detail to the voice recorder.]* ※*Q3 Please explain how "Modeling" was done in this section?* ※*Q4* Please explain the content of the ideal gas law.
Model Validation	***[model validation]:*** *Judging if the text description, chemical formula, and mathematical equations are reasonable, and determining whether factors that will affect theory can be ignored, judging the feasibility of experimental methods, and considering alternative solutions is called "model validation".*

[a]Non-italics in the Appendices indicate the content translated from the textbook, pages 52–56. Images and texts with permission from Han Lin Publishing Co., LTD.

Modeling Stage	Modeling-Based Text

[a]According to the ideal gas law and the relationship among gas pressure, volume, and temperature, there are limitations during use as shown in Figure 5-1a, b, c.

Figure 5-1a: Under constant temperature, when volume decreases, pressure would increase.

Based on PV = nRT; when the amount of gas and the temperature is fixed, PV = constant. When volume decreases, pressure will increase; when volume is increased, pressure will decrease.

Figure 5-1b: Constant volume, the temperature becomes higher, the pressure becomes larger.

Based on PV = nRT; the amount of gas and the volume is fixed, P/T = constant. The temperature becomes higher, the pressure becomes larger; the temperature becomes lower, the pressure becomes smaller.

Figure 5-1c: Under constant pressure, when temperature is raised, volume will increase.

Based on PV = nRT, under constant pressure and volume, V/T = constant. When the temperature is raised, volume would increase; when the temperature is lowered, volume would decrease.

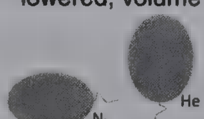


Figure 5-2: The helium (He) balloon floats since the density of He is lower than the density of air. The nitrogen (N_2) balloon sinks since the density of N_2 is similar to the density of air.

Based on PM = dRT, under constant temperature and pressure, the density of the gas is proportional to its molecular mass. Figure 5-2 shows the He balloon floating and the N_2 balloon lying on the table. This is because the molecular mass of He = 2, which means its density is lower than the density of N_2, whose molecular mass equals 28. In other words, the smaller the molecular weight (i.e., He), the lower the density; the greater the molecular weight (i.e., N_2), the greater the density.

[Explain the problem: Please express your thoughts in detail to the voice recorder.]

※Q5 How do you carry out "model validation"?

※Q6 Explain how the experiments of the ideal gas law and the formulae are consistent?

[a]Non-italics in the Appendices indicate the content translated from the textbook, pages 52–56. Images and texts with permission from Han Lin Publishing Co., LTD.

Modeling Stage	Modeling-Based Text
Model Analysis	***[model analysis]:*** *After the established model is validated, it is necessary to analyze the data to identify the relationship between the variables and evaluate the differences between predicted and experimental results. This is known as "model analysis".* [a]According to the equation PM = dRT, gas with lower density will have smaller molecular mass as well (Figure 5-2). Try to think about it: In Figure 5-2, if the volumes of the balloons filled with N_2 and He are the same, will their mass be the same as well? For PM = dRT, density (d) can be written as W / M, or $PV = \frac{W}{M}RT$. When the external conditions: pressure (P), temperature (T), and volume (V) are held constant, mass (W) will be proportional to molecular mass (M). The molecular mass of N_2 (M = 28) is greater than the molecular mass of He (M = 2). Consequently, for balloons with the same volume, the mass of the N_2-filled balloon would be greater. *[Explain the problem: Please express your thoughts in detail to the voice recorder.]* *⁎Q7 How can "model analysis" be carried out?* *⁎Q8 Analyze and evaluate whether the ideal gas law is reasonable based on the diagram provided above.*
Model Deployment	***[model deployment]:*** *The strengthening of the explanatory power of the model when the model is applied to other similar situations is called "model deployment".* [Example 5-1] The volume of a large helium balloon carrying large instruments was measured as 1.00×10^6 liters on the ground under the temperature of 27°C and pressure at 754 mmHg. If the balloon goes up to an altitude of 37 km, temperature would drop to −33°C and air pressure to 75.4 mmHg; what would be the balloon's volume in liters? [Answer] According to the ideal gas law PV = nRT, since the number of moles of the gas n and R was both constant, $PV \propto T \text{ or } \frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$ When on the ground: $P_1 = 754$ mmHg, $T_1 = (27+273)$ K = 300 K, $V_1 = 1.00 \times 10^6$ At high altitude: $P_2 = 75.4$ mmHg, $T_2 = (-33+273)$ K = 240 K, V_2 unknown Substituting the numbers into the above equations, we get the following: $V_2 = \frac{P_1V_1T_2}{P_2T_2} = \frac{(754\ mmHg)(1.00\times10^6\ liters)(240\ K)}{(75.4\ mmHg)(300K)} = 8.00 \times 10^6\text{liters}$ [Practice 5-1] Under STP, an ideal gas's volume is 400 ml; if the temperature goes up to 819°C, its volume changes to 200 ml. What is the pressure of the gas? [Answer] 8 atm [Example 5-2] Carbon monoxide (CO) is a colorless, odorless, highly toxic gas. It is produced when carbon-based material combusts with an insufficient oxygen supply. Under STP, the mass of the 400 ml of CO gas is 0.500 g. Assuming CO is an ideal gas, what is its molecular mass? [Answer] Under the STP condition, the gas pressure is 1atm with a temperature of 0°C = 273 K. According to the ideal gas law PM = dRT: $1\text{atm} \times M = (0.5/400) \times 0.082 \times 273;\ M = 35.7$ *[Explain the problem: Please express your thoughts in detail to the voice recorder.]* *Q9 How do you carry out "model deployment"?*
Model Reconstruction	***[model reconstruction]:*** *Thinking of its limitations and re-adjusting the model when the model cannot be applied to other situations is called "model reconstruction".*

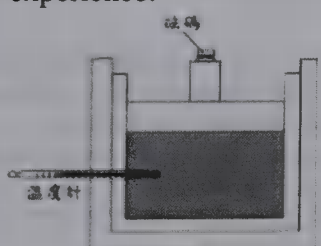
[a]Non-italics in the Appendices indicate the content translated from the textbook, pages 52–56. Images and texts with permission from Han Lin Publishing Co., LTD.

Modeling Stage	Modeling-Based Text
	[a]Do all gases fit the ideal gas law? If its features are fully consistent with the ideal gas equation, then it would be called an ideal gas. An ideal gas is a theoretical gas. From the submicroscopic point of view, it is assumed that the ideal gas molecules are very far apart and that the molecule itself has a mass but not volume. Also, there are no intermolecular forces to attract or repel the molecules. However, in reality, none of the gases fit the ideal gas description. Real gas molecules have both volume and intermolecular forces. Consequently, real gas and ideal gas would have varying degrees of differences under varied conditions. Deviations between the real gas and the ideal gas depend on the type of gas, temperature, and pressure. Experiments show a gas under lower temperature or higher pressure behaves most closely to an ideal gas. For example, helium has the lowest boiling point of all gases, 4 K. Its intermolecular attraction is small, so helium is the closest real gas to the ideal gas. Conversely, the intermolecular forces of ammonia, carbon dioxide, and chlorine gas are much greater, as such the difference between them and the ideal gas is much more significant. However, in most cases, such differences remain negligible. Therefore, if the gas particles fit the ideal gas law-$PV = nRT$'$PM = dRT$, other detailed condition equations (including: Boyle's law, Charles and Gay-Lussac's law, and Avogadro's law) will also conform to the characteristics of an ideal gas. *[Explain the problem: Please express your thoughts in detail to the voice recorder.]* *⁎Q10 How can "model reconstruction" be carried out?* *⁎Q11 Why does a gas behave more like the ideal gas under lower temperature or higher pressure?* *⁎Q12 Did you correct any of your original concepts during the process of reading this section? Please provide details on your concepts and the process of how your concepts changed.*

[a]Non-italics in the Appendices indicate the content translated from the textbook, pages 52–56. Images and texts with permission from Han Lin Publishing Co., LTD.

APPENDIX B: SELECTED ITEMS IN THE GAS PARTICLE CONCEPTS ASSESSMENT

I. The airtight container in the drawing is filled with a proper amount of nitrogen. The piston above the container is movable. We can control the volume of the container by using weights and make temperature adjustments (i.e., increasing or decreasing the temperature of the gas in the container) with a device that provides uniform heating. Answer the following question sets based on the above description and your own experience:



1.1 What do you think causes the gas pressure in the airtight container in the drawing?
   (A) Weight of the weights.
   (B) Weight of the weights and atmospheric pressure.
   (C) The gas particles in the container pressing against one another.
   (D) The gas particles in the container pressing against the container wall.
   (E) Collision of fast-moving gas particles in the container.

(F) Fast-moving gas particles in the container hitting the container wall continuously.
(G) Other:__________

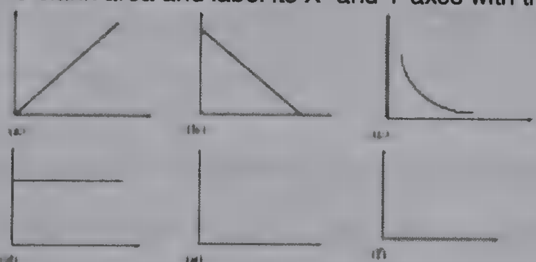
1.2 Given the same temperature and volume, what do you think is the main factor that influences the gas pressure in the container?
(A) Total weight of the gas.
(B) Sizes of molecules.
(C) Activity.
(D) Attraction between gas particles.
(E) Number of molecules.
(F) Energy loss due to collision of gas particles.
(G) Related to volume of the container only.
(H) Related to external temperature only.
(I) Other:_________

1.3 If the volume of the airtight container is compressed to half of its original size, with both the temperature and the number of molecules unchanged, how will the gas pressure in the container change in comparison with the gas pressure before the compression?
(A) Falls.
(B) Remains unchanged.
(C) Rises.
(D) No idea.

1.4 Continued from 1.3, what do you think causes the gas pressure in the airtight container to change?
(A) The gas particles in the container press more against one another once the volume of the container is reduced.
(B) The total weight of the gas is reduced.
(C) Activity of the gas.
(D) A change in the attraction between gas particles.
(E) Weight of the weights.
(F) The gas particles in the container collide more frequently with one another once the space in the container is reduced.
(G) Energy loss due to collision of gas particles.
(H) Other:__________

APPENDIX C: SELECTED SEMISTRUCTURED INTERVIEW QUESTIONS ON MODELING COMPETENCIES

	Referring to the drawing on the left, when a diver 20 m below sea level exhales in the water, the gas exhaled forms ascending bubbles. Answer the following questions by applying the scientific knowledge you have learned or simply based on common sense.
Modeling Process	Questions on the "Ideal Gas Law"
Model Selection	Q1. Do the bubbles increase in size, decrease in size, or remain the same size while ascending toward the sea surface? Q1-1. If the bubbles change in size, what are the possible variables? Q2. Is there anything to be found inside the bubbles if a super powerful electron submicroscope can be used to observe the inside of the bubbles? If yes, try to draw what is to be found in the bubbles.

<table>
<tr><th>Modeling Process</th><th>Questions on the "Ideal Gas Law"</th></tr>
<tr><td></td><td>Q2-1. If nothing is to be found in the bubbles, try to explain the composition of the bubbles.
Q3. If there is something to be found in the bubbles, what is the shortest distance (gap) between such things in a given bubble? And why?
Q4. Define the variables you mentioned in Q1-1 from the perspective of gas particles.</td></tr>
<tr><td>Model Construction</td><td>Q5. If you are asked to describe the phenomenon of bubble size variation taking place while the bubbles ascend toward the sea surface, how would you describe the relationship between the variables in Q1-1? Please elaborate.
Q6. From the perspective of gas particles, explain the variable relationship you described in Q5.
Q6-1. What is the condition inside the bubbles when they are approaching the sea surface and are about to burst?</td></tr>
<tr><td>Model Validation</td><td>Q7. Design an experiment to demonstrate that the variable relationship you described in Q5 is reasonable.
Q8. From the perspective of gas particles, explain the variable relationship as observed in the experiment of Q7.</td></tr>
<tr><td>Model Analysis</td><td>Q9. Based on the variable relationship you described in Q5, label the X- and Y-axes of the applicable X-Y graph below with the possible variables. If none of the graphs below reflect the relationship you described, plot your own graph in the blank area and label its X- and Y-axes with the variables.
</td></tr>
<tr><td>Model Analysis</td><td>Q10. Based on the X-Y graph you chose or plotted in Q9, give the equation of the variable relationship.
Q11. Based on the table below, which is assumed to show data actually collected, determine whether the relationship/equation you chose/gave in Q9 and Q10 is reasonable. Elaborate on your reasons.
※ The table below shows data actually collected, with the following assumptions: the balloon is filled with helium; temperature decreases by 0.6°C for every 100-m increase in elevation; the atmospheric pressure around ground level decreases by about 8 mmHg (0.8 cm-Hg) for every 100-m increase in elevation; and the balloon does not burst during the process.
<table>
<tr><td>Elevation (m)</td><td>0.0</td><td>50.0</td><td>100.0</td><td>500.0</td><td>1000.0</td></tr>
<tr><td>Balloon Size (mL)</td><td>3000.0</td><td>3012.8</td><td>3025.7</td><td>3134.3</td><td>3284.3</td></tr>
<tr><td>Air Temperature (°C)</td><td>20.0</td><td>19.7</td><td>19.4</td><td>17.0</td><td>14.0</td></tr>
</table></td></tr>
<tr><td>Model Deployment</td><td>Q12. The chemistry teacher produces hydrogen by reacting a magnesium ribbon with hydrochloric acid, and the hydrogen is collected and injected into balloons. During the injection process, one of the balloons gets loose from the operator's hand and ends up mixed with some helium balloons already blown. Assume there are two balloons that are blown to the same size with different gases, respectively (say, hydrogen and helium) and fly away from the same spot. Based on your theory in Q5, describe how these two balloons would fly and explain why.
Q13. Explain your description in Q12 from the perspective of gas particles.</td></tr>
</table>

Modeling Process	Questions on the "Ideal Gas Law"
Model Reconstruction	Q14. John's father drives his car to a workshop for maintenance, and one of the tires is replaced. To begin with, the old tire is removed. Then, a new tire is mounted and inflated until a stable tire pressure is achieved. How would you describe the variables in the tire inflation process, the relationship between the variables, and a plot showing the relationship? Q15. From the perspective of gas particles, explain the variables and the relationship you described in Q14. Q16. Continued from Q14, after the tire replacement, John's father drives the entire family down the highway to Kenting in search of shooting locations for the movie Cape No. 7. How would you describe the variables that influence the tires of the car during the long hours in which the car runs at a high speed? How would you describe the relationship between the variables and a plot showing the relationship? Q17. From the perspective of gas particles, explain the variables and the relationship you described in Q16. Q18. Continued from Q16, after quite a long drive, John and his family finally arrive at their hotel in Kenting and begin a series of activities as soon as the car is parked. How would you describe the variables that influence the tires after the car has stopped and stays parked for some time? How would you describe the relationship between the variables and a plot showing the relationship? Q19. From the perspective of gas particles, explain the variables and the relationship you described in Q18.

REFERENCES

Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy. New York: Academic Press.

Blank, L. M. (2000). A metacognitive learning cycle: A better warranty for student understanding. Science Education, 84(4), 486–506.

Branbrand, C., & Dahl, B. (2009). Using the SOLO taxonomy to analyze competence progression of university science curricula. Higher Education 58(4), 531–549.

Chambers, K. S., & Andre, T. (1997). Gender, prior knowledge, interest and experience in electricity and conceptual change text manipulations in learning about direct current. Journal of Research in Science Teaching, 34(2), 107–123.

Chang, C. K., & Chiu, M. H. (2009). The development and application of modeling ability analytic index—take electrochemistry as an example. Chinese Journal of Science Education, 17(4), 319–342.

Chi, M. T. H., Leeuw, N, Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18(3), 439–477.

Chiu, M. H. (2012). Via the use of web-based diagnosis system and interactive learning environment investigating students mental models of particles and their modeling ability (Technical report). Taipei, Taiwan: National Science Council.

Clement, J. (1989). Learning via model construction and criticism. In G. Glover, R. Ronning, & C. Reynolds (Eds.), Handbook of creativity, assessment, theory and research (pp. 341–381). New York: Plenum Press.

Clement, J. (2000). Model based learning as a key research area for science education. International Journal of Science Education, 22(9), 1041–1053.

Clement, J. (2008). Student/teacher co-construction of visualizable models in large group discussion. In J. Clement & M. A. Rea-Ramirez (Eds.), Model based learning and instruction in science (pp. 11–22). Dordrecht, The Netherlands: Springer.

DeBerg, K. C. (1989). The emergence of quantification in the pressure-volume relationship for gases: A textbook analysis. Science Education, 73(2), 115–134.

Dori, Y. J., & Kaberman, Z. (2012). Assessing high school chemistry students' modeling sub-skills in a computerized molecular modeling learning environment. Instructional Science, 40(1), 69–91.

Gericke, N. M., & Hagberg, M. (2010). Conceptual incoherence as a result of the use of multiple history models in school textbooks. Research in Science Education, 40(4), 605–623.

Gericke, N. M., Hagberg, M., & Jorde, M. (2013). Upper secondary students' understanding of the use of multiple models in biology textbooks—The importance of conceptual variation and incommensurability. Research in Science Education, 43(2), 755–780.

Giere, R. N. (1988). Explaining science: A cognitive approach. Chicago: University of Chicago.
Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), Visualization in science education (pp. 1–27). Dordrecht, The Netherlands: Springer.
Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. Journal of Science Education and Technology, 13(1), 7–22.
Grosslight, L., Unger, C., Jay, E., & Smith, C. (1991). Understanding models and their use in science conceptions of middle and high school students and experts. Journal of Research in Science Teaching, 28(9), 799–822.
Halloun, I. (1996). Schematic modeling for meaningful learning of physics. Journal of Research in Science Teaching, 33(9), 1019–1041.
Halloun, I. (2004). Modeling theory in science education. Boston: Kluwer.
Hestenes, D. (1992). Modeling games in the Newtonian world. American Journal of Physics, 60(8), 732–748.
Hestenes, D. (1995). Modeling software for learning and doing physics. In C. Bernardini, C. Tarsitani, & M. Vincentini (Eds.), Thinking physics for teaching. (pp. 25–66). New York: Plenum Press.
Hestenes, D. (2010). Modeling theory for math and science education. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), Modeling students' mathematical modeling competencies. (pp. 13–41). New York: Springer.
Justi, R. S., & Gilbert, J. K. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. International Journal of Science Education, 24(12), 1273–1292.
Kaberman, Z., & Dori, Y. J. (2009). Metacognition in chemical education: Question posing in the case based computerized learning environment. Instructional Science, 37(5), 403–436.
Kautz, C. H., Heron, P. R. L., Loverude, M. E., & McDermott, L. C. (2005a). Student understanding of the ideal gas law, part I: A macroscopic perspective. American Journal of Physics, 73(11), 1055–1063.
Kautz, C. H., Heron, P. R. L., Loverude, M. E., & McDermott, L. C. (2005b). Student understanding of the ideal gas law, part II: A microscopic perspective. American Journal of Physics, 73(11), 1064–1071.
Kennedy, A. G. (2012). A non representationalist view of model explanation. Studies in History and Philosophy of Science Part A, 43(2), 326–332.
Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. Studies in History and Philosophy of Science Part A, 42(2), 262–271.
Lesh, R., & Doerr, H. M. (2003). Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching. Mahwah, NJ: Erlbaum.
Levy, S. T., & Wilensky, U. (2009). Students' learning with the connected chemistry (CC1) curriculum: Navigating the complexities of the particulate world. Journal of Science Education and Technology, 18(3), 243–254.
Lin, H., Cheng, H., & Lawrenz, F. (2000). The assessment of students' and teachers' understanding of gas laws. Journal of Chemical Education, 77(2), 235–238.
Liu, C. K., & Chiu, M. H. (2010). From modeling perspectives to analyze modeling processes of atomic theory in senior high school chemistry textbooks and their implications. Research and Development in Science Education Quarterly, 59, 23–54.
Minogue, J., & Jones, G. (2009). Measuring the impact of haptic feedback using the SOLO taxonomy. International Journal of Science Education, 31(10), 1359–1378.
Morgan, M. S., & Morrison, M. C. (1999). Models as mediating instruments. In M. S. Morgan & M. C. Morrison (Eds.), Models as mediators (pp. 10–37). Cambridge, England: Cambridge University Press.
Niaz, M. (1995). Relationship between student performance on conceptual and computational problems of chemical equilibrium. International Journal of Science Education, 17(3), 343–355.
Niaz, M., & Robinson, W. R. (1992). From algorithmic mode to conceptual gestalt in understanding the behavior of gases: An epistemological perspective. Research in Science & Technological Education, 10(1), 53–64.
Palmer, D. H. (2003). Investigating the relationship between refutational text and conceptual change. Science Education, 87(5), 663–684.
Saari, H., & Viiri, F. (2003). A research-based teaching sequence for teaching the concept of modelling to seventh-grade students. International Journal of Science Education, 25(11), 1333–1352.
Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Designing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. Journal for Research in Science Teaching, 46(6), 632–654.
Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. Cognition & Instruction, 23(2), 165–205.
Smith, M. U., & Adkison, L. R. (2010). Updating the model definition of the gene in the modern genomic era with implications for instruction. Science & Education, 19(1), 1–20.

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. International Journal of Science and Mathematics Education, 8(6), 951–970.

Wang, C. Y., & Barrow, L. H. (2011). Characteristics and levels of sophistication: An analysis of chemistry students' ability to think with mental models. Research in Science Education, 41(4), 561–586.

Zöttl, L., Ufer, S., & Reiss, K. (2011). Assessing modelling competencies using a multidimensional IRT approach. In G. Kaiser, W. Blum, R. B. Ferri, & G. Stillman (Eds.), Trends in teaching and learning of mathematical modelling (pp. 427–437). Ne w York: Springer.

Scientific Literacy for Participation: A Systemic Functional Approach to Analysis of School Science Discourses, by Erik Knain. Sense Publishers, Rotterdam, The Netherlands, 2015. xv + 165 pp. ISBN 978-94-6209-894-7.

Science is a global human endeavor involving the use of disciplinary methods to investigate the natural phenomenon as well as texts to construct and communicate processes and understandings. This conception of science reaffirms the traditional commitment to practical activities (i.e., hands-on manipulation of the material world) but more importantly emphasizes the centrality of language-based work (i.e., reading, writing, and argument) to science education. It has stimulated a fruitful line of international scholarship that promotes interaction between language/literacy and science. An example of this synergy is Erik Knain's *Scientific Literacy for Participation* (SLP).

SLP draws its theoretical framework and analytical tools from systemic functional linguistics (SFL) (Halliday & Matthiessen, 2004), an approach to linguistics that considers language as a social semiotic system linking the structural and functional units of language to their meaning-making potential in particular contexts of use. It provides a basic introduction to some of the key constructs in functional grammar (e.g., *text*, *context*, *genre*, *metafunction*), describes some key lexicogrammatical and visual resources (e.g., *nominalization*, *conjunction*, *modality*, *image*, *gesture*) that are important to scientific meaning making, and exemplifies several linguistic tools (e.g., *process analysis, cohesion analysis*, *modality analysis*) for analyzing the content, organization, and perspective of science texts from pedagogical contexts. The book recognizes the essential role of texts (oral and written) in science education and demonstrates the power of language analysis in understanding the nature of science, assessing students' science learning, and reconceptualizing science curricula and pedagogy.

An underlying message of SLP is that science depends on not only verbal resources but also visual resources such as mathematical symbols, gestures, and images to make meaning. Each resource system has its own affordances and limitations and must be integrated with others to be of use to science. Together, the verbal and visual resources enable scientists to present information, infuse ideology, and structure discourse in ways that conform to disciplinary norms and connect with the broader social, political, and economic institutions in which they operate. Thus, as SLP suggests, facility in understanding and using both resources should be an essential component of science literacy.

Another message underpinning the book is that texts are motivated in the sense that the verbal and visual choices made by the author "are designed to convey particular meanings in particular ways and to have particular effects" (Janks, 2010, p, 153). This suggests that text meaning is neither natural nor neutral. From this perspective, science does not consist of an immutable body of "facts"; rather, it evolves in response to new technologies and

discoveries as well as to the sociopolitical and economic contexts that make its continued existence possible. This character of science necessitates, as SLP argues, the use of critical discourse analysis, an approach to reading that promotes thoughtful analysis of text and interrogation of the ideologies underpinning the text. The approach requires a deep understanding of how language and other semiotic choices realize meaning in text. Explicit attention to how texts mean (e.g., the selection, juxtapositioning, sequencing, and layout of semiotic choices) can yield important insights into the meaning of the text, uncover evidence about students' learning, and enable participation in decision-making processes involving science, technology, and society. Developing a range of tools for critically analyzing and appraising the language and visual representations in text, therefore, promises more robust development of science literacy.

The book, written for graduate students in science education, aims to "offer an in-depth understanding of language, meaning, and text in school science" (p. ix). It falls short of this goal, however, due to a number of major weaknesses. The first weakness concerns the conceptual organization of the book and some of its chapters. For a book that purports to showcase the relevance of language and the power of language analysis to science education, I expect it to begin with a chapter that conceptualizes science as a discipline where language and other semiotic resources are integral to its social practices. Subsequent chapters will then (a) describe a theoretical framework (such as SFL) that is powerful for making sense of language and visual representations and their meaning making potential in science, (b) illustrate how SFL tools can be applied in classroom discourse analysis and the reading/writing of scientific texts, and (c) discuss the implications of these analyses for science teaching/learning and perhaps even science teacher education. While most of these threads can be gleaned from the book, they are not coherently organized. For example, chapter 1, misleadingly titled "An Anatomy of Discourses," explains some key SFL constructs (e.g., *text*, *context*, *genre*, *metafunction*) that seem better suited for chapter 2, where Halliday's functional grammar is described. Chapters 3–6 could perhaps be restructured to more logically and clearly show (a) how SFL tools can be used to analyze texts from different pedagogical contexts (e.g., classroom interaction, texts students are expected to read, texts students write, official curriculum documents), (b) which tools are most relevant to and particularly revealing for which type of text or analytical purpose, and (c) what these analyses can reveal in terms of content, structure, identity, value, agency, power relations, language skills, literacy challenges, or other critical issues germane to science education.

The same issue with organization also exists in individual chapters. It is often unclear how different sections within a chapter relate to one another and to the overall focus of the chapter. For example, the various sections in chapter 1 appear disjoined; they are supposed to be linked by a conceptual framework (i.e., functional model of language) that has yet to be clearly and concisely articulated. In chapter 2, the last two sections (pp. 43–57) seem rather loose and conceptually detached from the rest of the chapter. In some chapters, concepts and models outside SFL, such as James Gee's big D and small d (chapter 2) and Norman Fairclough's critical discourse analysis (chapter 6), are introduced, but their relations to SFL and the chapter focus, as well as their implications for the design and analysis of learning activities, are not clearly explicated.

The second major weakness of the book has to do with the accuracy and rigor of its content. In many places throughout the book, there are inaccuracies, inconsistencies, gaps, errors, and confusions in the interpretation and application of SFL and related work. In chapter 1, for example, genre is described as "an intermediate level between the act of meaning in the situation and a cultural level" (p. 9). This conception is not accurate. In SFL, genre is concerned with how a text is organized to achieve its social purpose and as

such, it realizes the context of culture. Register (not mentioned in the book), on the other hand, operates at the level of the specific situation within the culture; it realizes genre and is realized in language choices. There are also problems with text analyses involving, for example, parsing of clauses, classification of process types, and identification of clause relations.

A further issue with content is the omission of key information. The discussion of ideational meaning in chapter 2 focuses on the experiential (how language organizes experience), but ignores the logical (logical-semantic relations among clauses). The discussion of interpersonal meaning, and subsequently of critical discourse analysis (chapter 6), does not draw on appraisal theory (Martin & White, 2005), a more recent development in SFL. Analysis of mood choices (i.e., resources for making statements, asking questions, giving commands, and making offers) is important for oral interaction, but not particularly revealing for understanding ideology in written texts. For this reason, a focus is warranted on resources (besides modality) for expressing judgment and emotion (attitude), for upgrading or toning down evaluation (graduation), and for building the "authorial self" (engagement). In discussing textual meaning, the notions of Given and New are introduced, but the analysis of textual meaning in later chapters focuses on the notions of Theme and Rheme. Although the two are related, they are not the same. Given and New are terms for information structure; they are speaker oriented and indicated through intonation. Theme and Rheme are terms for thematic structure; they are speaker oriented and signaled by position of clause. The conflation of the two, as well as the inconsistencies, is likely to confuse readers new to SFL.

Another major weakness of the book is the lack of writing fluency. Many sentences throughout the book are worded awkwardly. There are also issues with spelling, grammar (e.g., subject–verb agreement, preposition usage, pronominal reference), punctuation, and repetition that reduce clarity of message, disrupt discursive flow, and engender confusion. Moreover, citations do not follow a consistent format, and not all cited works appear in the References section. Some quotes do not seem to align well with the messages they are supposed to support. Space limitation precludes an inventory of examples here. Suffice it to say that these discursive problems make the book less accessible to the audience for which it intends.

As a language and literacy education scholar with an interest in science education, I was initially excited at the request to review SLP. However, my enthusiasm faded as I delved further into the book. Issues with the book's organization, content, and style obscure its sometimes important messages, overshadow some insightful analyses (with respect to, for example, the energy transfer text in chapter 4), sabotage the goals it sets out to accomplish, diminish its credibility, and undermine the reader's confidence in what could have been a valuable contribution to the scholarship on science literacy.

REFERENCES

Halliday, M., & Matthiessen, C. (2004). An introduction to functional grammar (3rd ed.). London: Routledge.

Janks, H. (2010). "Language as a system of meaning potential": The reading and design of verbal texts. In T. Locke (Ed.), Beyond the grammar wars (pp. 151–169). London: Routledge.

Martin, J., & White, P. (2005). The language of evaluation. New York: Palgrave Macmillan.

ZHIHUI FANG
School of Teaching and Learning
University of Florida
Gainesville, FL 32611–7048, USA

DOI 10.1002/sce.21178
Published online 6 August 2015 in Wiley Online Library (wileyonlinelibrary.com).